

Ideas in Dialogue: The Effects of Interaction on Creative Problem Solving

Christine Howes (christine.howes@gu.se)

University of Gothenburg
Centre for Language Technology, Sweden

Patrick G.T. Healey (p.healey@qmul.ac.uk)

Queen Mary University of London
Cognitive Science Research Group, UK

Pietro Panzarasa (p.panzarasa@qmul.ac.uk)

Queen Mary University of London
School of Business and Management, UK

Thomas Hills (thomhills@gmail.com)

University of Warwick
Department of Psychology, UK

Abstract

Much problem-solving research has investigated if and why ‘two heads are better than one’, but typically posits that if there is any process gain observed it is because of the exposure to the ideas provided by another person’s attempted solutions. This work fails to acknowledge or investigate what the interaction itself contributes to joint problem solving.

Using an online version of the *Alternative Uses Task*, we compare situations in which people are passively exposed to what is said in a dialogue with situations in which people are actively engaged in the dialogue, thus varying the interactivity independently of the informational content that participants were exposed to.

Interacting participants produce more turns overall, but they do not come up with more ideas. Interacting participants were also more likely to build on each other’s ideas and produce more complex ideas when a turn is linked to a previous idea; following leads to elaboration – but only if there is genuine interactivity. These results indicate that conversational mechanisms promote the exploration of a problem space and that merely counting the number of ideas produced would miss the importance of the interaction itself.

Keywords: interaction; dialogue; creative problem-solving

Introduction

Much problem-solving research has investigated if and why ‘two heads are better than one’ (Hill, 1982), with mixed results. In brain-storming and other creative thinking studies, a large body of work has found that pooled individuals (*nominal groups*) out-perform real groups of the same number of people, with several reasons suggested for this productivity gap including social loafing, performance matching and evaluation apprehension (Mullen, Johnson, & Salas, 1991). However, statistical pooling of non-interacting individuals means this is not a fair comparison; in groups, participants can not produce ideas simultaneously (*production blocking*), so that over the same time course individuals have more opportunities to present ideas because they do not have to wait whilst listening to others’ contributions (Kerr & Tindale, 2004).

The constraint of production blocking has been alleviated with the advent of computer-mediated communication (Dennis & Williams, 2003), and, in a non-interactive online paradigm, individuals exposed to others’ ideas performed better than those who were not (Nijstad, Stroebe, & Lodewijckx, 2002). This work posits that it is exposure to the ideas provided by another’s attempted solutions that provides process gain, thus offering a cognitive perspective. However, social aspects have also been found to influence group performance. For example, where groups were presented with the same

ideas from apparently different sources, there were greater gains when they were told that the information came from a previous participant, rather than being randomly generated by a computer (Dugosh & Paulus, 2005). Similarly, studies have found that even where nominal groups seem to out-perform interacting groups, this depends on the assessment criteria used; for example, Kohn, Paulus, and Choi (2011) found that groups generated more novel combinations than nominal groups when combining rare ideas.

Previous studies have ignored how the process of interaction may itself shape the production of problem solutions. For example, Ziegler, Diehl, and Zijlstra (2000) found no improvement for interacting groups, who produced more ‘irrelevant utterances’. However, conversation involves more than just exposure to what another person says; it is built up from sequences of collaborative contributions that directly build on each other using mechanisms that are specific to interaction (Sacks, Schegloff, & Jefferson, 1974; Clark, 1996; Goodwin, 1979). This has two potentially important consequences for group problem solving. Firstly, it is problematic to count ‘ideas’ as independent events; one person’s idea will only be properly understood in the context of what other people say in the preceding and succeeding turns. Secondly, the mechanisms of interaction – such as establishing joint reference, articulating problem constraints or clarification and repair – will themselves shape the products of idea generation in addition to individual cognitive factors.

In order to test the impact of these factors we therefore need to compare situations in which people are passively exposed to what is said in a dialogue with situations in which people are actively engaged in the dialogue. We present a novel experimental set-up to investigate the input of interaction on creative problem solving, whilst matching the informational content and timing of ideas received by participants in an *interactive* and a non-interactive *playback* condition (see Methods section, below, for details).

Hypotheses

Following Nijstad et al. (2002), we hypothesise that in a creative idea-generating task participants exposed to information content from others should perform better than those not exposed to the additional content. However, we also hypothesise that exposure to others’ ideas in an interactional setting should have additional benefits, independently of the informational content. Specifically:

1. Non-interacting dyads should come up with more or more complex ideas than individuals not exposed to others' ideas
2. Interacting dyads should come up with more or more complex ideas than non-interacting dyads or individuals
3. Interacting dyads should build on each others' ideas more than non-interacting dyads who are exposed to the same ideas

Method

The DiET chat tool

The Dialogue Experimental Toolkit (DiET) chat tool is a text-based chat interface into which interventions can be introduced into a dialogue in real time (Healey, Purver, King, Ginzburg, & Mills, 2003). As these manipulations occur as the dialogue progresses, they cause minimal disruption to the 'flow' of the conversation.

The DiET chat tool is a custom built Java application, consisting of two main components: the server console and the user interface. The server time-stamps and stores each key press, and acts as an intermediary between what participants type and what they see. Each turn is passed to the server, from where it is relayed to the other participants.

User interface The user interface is designed to look and feel like common instant messaging applications (see Figure 1). It consists of a display split into two windows, with a status bar between them. The ongoing dialogue, consisting of both the nickname of the contributor and their transmitted text, is shown in the upper window. In the lower window, participants can type and revise their contributions, before sending them to their co-participants. All key presses are time-stamped and stored by the server. The status bar, between the upper and lower windows, shows whether any participants are actively typing.

The conditions For the current experiment, two windows were visible on each participant's screen (see Figure 1): a) the instruction window, which showed instructions about the task, and details of the current item, and b) the user interface or chat window, as discussed above.

Participants were recruited in pairs and assigned to one of three conditions; i) *interactive* ii) *playback* iii) *individual*. In the *interactive* condition, participants conversed with each other and came up with solutions to the task together (see e.g. (1)). In the *playback* condition, participants came up with solutions on their own, however, they were also presented with the contentful suggestions made by a single participant from a dyad in the interactive condition in their chat window (see e.g. (2)-(3)). The timings from the original conversation were preserved, thus these participants were exposed to the same ideas, at the same pace, as the genuine partner. In the *individual* condition the participant completed the task alone. *Nominal pairs* were subsequently created by interweaving the transcripts of turns by two participants in the *individual* condition, who did not interact and did not see anybody else's

suggestions, with the timings of each individual's contributions preserved (see e.g. (4)). This manipulation allows us to independently vary the informational content and the interactivity that participants were exposed to.

Subjects

68 English speaking students were recruited for the experiment, in pairs. All had previous experience using internet chat software. Each subject was paid £7 for their participation.

Task

Following Gilhooly, Fioratou, Anthony, and Wynn (2007), we used the Alternative Uses Task, which is a common task for assessing creativity. Participants were presented with the instructions as shown in Figure 1.

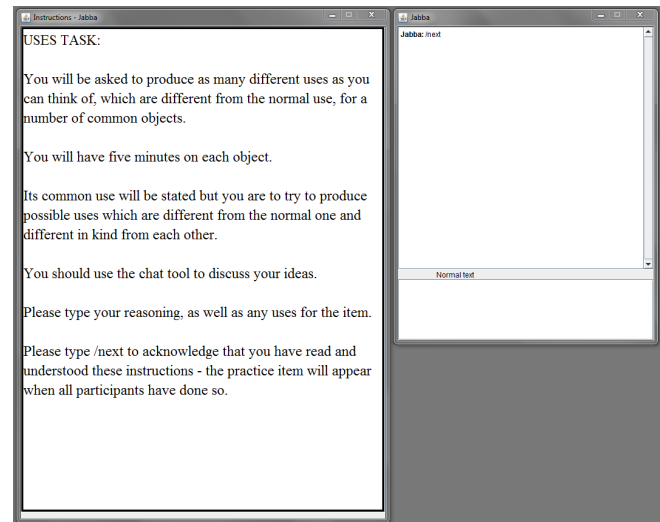


Figure 1: The instructions window and DiET chat window

The practice item was *newspaper*. After five minutes had elapsed, participants were alerted to the time being up, and shown the following text:

The common use for a newspaper is for reading.

Possible other uses include:

- ... for swatting flies...
- ... to line drawers...
- ... to make a paper hat...
- ... and so on.

If you have any questions about this task, please ask the experimenter now.

Otherwise the test will begin when all of you have typed /next

The test items were *barrel*, *brick*, *car tyre* and *pencil*. These were presented in a random order, and after five minutes on each item, participants received a message in both their chat and instructions windows to alert them to the change of item.

Tag	Value	Explanation	kappa
is-use	y/n	For all turns: is this turn a suggested use for the item?	0.86
continuesYN	y/n	Where is-use = y: does this turn develop or repeat a previous suggestion?	0.70
continues	sentence ID	If so, which one?	0.68
similarity	1-5	For continuations: how similar is it to that previous suggestion?	0.62
complexity	1-5	For turns where is-use = y: how complicated/elaborate is the suggestion?	0.83

Table 1: Annotation Tags

Annotation scheme

Typed transcripts were annotated (using SCoRE; Purver, 2001) for a number of factors commonly used in assessing the Alternative Uses Task (Kaufman, Plucker, & Baer, 2008; Silvia, 2011), as shown in Table 1. For the *interactive* condition, these were direct transcripts of the interaction; in the *playback* condition, the transcripts include the replayed turns seen by participants and for the *individual (nominal pairs)* these are the constructed transcripts of the combined responses of two individuals. Excerpts of transcripts from the three conditions (on the item *brick*) are shown in (1)-(4).¹

- (1) Interactive dyad (C & N)
 - C: as a paper-weight
 - N: nice. You could use a lot of them to play Jenga
 - C: dangerous Jenga
 - N: I was going to follow with that
- (2) Playback individual (E, who saw C’s ideas – as in (1))
 - S: as a paper-weight
 - S: dangerous Jenga
 - E: door stop
- (3) Playback individual (J, who saw N’s ideas – as in (1))
 - S: You could use a lot of them to play Jenga
 - J: perhaps, though I feel that could be a health and safety risk
 - J: (they’re quite heavy and may cause damage upon impact)
- (4) Nominal pair (neither B nor M saw each other’s turns)
 - B: making a tower
 - M: shattered into smaller pieces and stuck on top of low walls to deter people getting into the premises
 - B: playing jenga

Annotators were asked to identify if a turn (a line of text entry in the chat tool) suggested a use for the item. For each turn that was judged to be a use (is-use=y), the turn was annotated for whether it developed or repeated a previously suggested use (continuesYN and continues), and if so how similar it was to that prior turn on a scale of 1-5. A similarity value of 1 indicates a direct repetition (for example, people often recycled ideas verbatim, such as ‘weapon’) and 5 indicates that the suggested use was mostly different

¹S: denotes server generated ‘playback’ turns

(e.g. for a barrel; ‘this you can use for floating and a raft’ was later followed by ‘transporting dwarfs down a river’, rated as similarity=4). Independently of this, annotators also gave each usage turn a complexity value between 1 and 5. In uses for a barrel, for example, ‘storing beer’ was rated as complexity=1, whilst ‘paint green and pretend it’s a shell for a tortoise’ was rated as complexity=5. This measure is intended to get at the notion of creativity; unlike many studies, we did not use a direct measure of uniqueness for this, as this is known to be sensitive to sample size (Silvia, Martin, & Nusbaum, 2009).

Two annotators both annotated a subset of the transcripts, and Cohen’s kappa values were calculated. The kappa statistic for each annotation tag, as well as the annotator instructions can be seen in Table 1.²

Analyses

Analyses were run using Generalised Linear Mixed Models to control for both fixed and random effects. In all reported models, participant and conversation were entered as random effects, and condition as a fixed effect. Models with a binary dependent variable (is-use; continuesYN) use a binomial distribution and logit link function, while models with a numerical dependent variable (similarity; complexity) use a gamma distribution with a log link function. We report exact p-values throughout, but take $p < 0.05$ to be the criterion of significance.

Results

As can be seen from Table 2, *interactive* pairs produce proportionally fewer turns that describe a possible use for the object than participants in the *playback* and *individual* conditions. A GLMM with is-use as dependent variable (DV) found a statistically significant main effect of condition ($F_{2,3599} = 12.053, p < 0.001$). Pairwise comparisons showed that the *interactive* pairs differed significantly from both the other conditions (see Figure 2;

²The kappa scores for the ordinal variables of similarity and complexity are weighted based on the squared distance between categories, such that disagreements involving distant values are weighted more heavily than disagreements involving more similar values (Agresti, 2002; Fleiss & Cohen, 1973). Similarity measures have the lowest kappa, however this figure is skewed by the cases which one annotator thought followed another turn (and therefore assigned a similarity measure) and the other did not (and therefore assigned no similarity measure) – with these cases removed, weighted kappas rise to 0.91.

Condition	Total turns					Average turns per item	
	is-use=y	%	is-use=n	%	total	is-use=y	total
Interactive	1078	58.9	753	41.1	1831	19.60	33.29
Playback	1192	88.3	158	11.7	1350	19.90	22.25
Individual	393	93.3	28	6.7	421	16.38	17.58

Table 2: Number of turns that constitute a use

Condition	continuesYN=n		continuesYN=y				total
		%	between	%	within	%	
Interactive	563	52.2	63	5.8	452	41.9	1078
Playback	727	61.0	99	8.3	366	30.7	1192
Nominal pair	246	62.6	50	12.7	97	24.7	393

Table 3: Idea turns that continue a prior turn within or between items

interactive/playback $t_{1,3599} = -3.682, p < 0.001$; *interactive/nominal pairs* $t_{1,3599} = -3.595, p < 0.001$).

However, these differences do not reflect poorer task performance; there is no difference in the average number of uses per “dialogue” between the conditions (see Table 2; $F_{2,188} = 0.718, p = 0.489$). Participants in the *interactive* condition used more turns per item than in the *playback* or *individual* conditions. Although *interactive* dyads produce more turns in total, they do not come up with more (or fewer) uses – nearly half of their turns are not presenting ideas, but instead are e.g. offering feedback to their interlocutor.

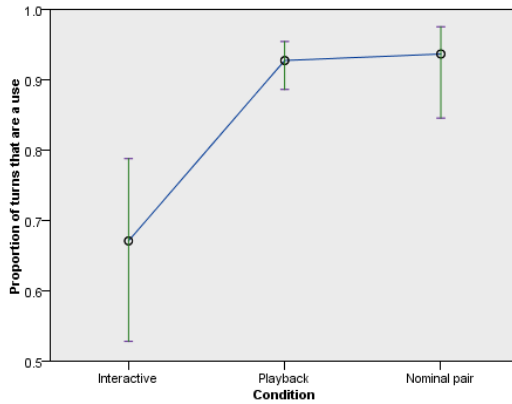


Figure 2: Marginal means of proportion of turns that are a use

In addition to the dialogue coordination turns, interactivity also changes the nature of the responses. As can be seen from Figure 3, *interactive* dyads idea turns are more likely to follow on from a previous turn in the conversation than those in the *nominal pairs* condition, and those in the *playback* condition, despite receiving the same informational content; *playback* dyads are not significantly different to the *nominal pairs* in this regard (GLMM with *continuesYN* as DV; $F_{2,2660} = 3.168, p = 0.042$; pairwise effects *interactive/playback* $t_{1,2660} = 2.160, p = 0.031$; *interactive/nominal pairs* $t_{1,2660} = 2.128, p = 0.033$).

In terms of complexity, *interactive* dyads don’t produce

significantly more complex ideas in general (see Figure 4; GLMM (DV *complexity*) – no main effect of condition; $F_{2,2660} = 1.169, p = 0.311$). However, they do produce more complex ideas than those in the *individual* and *playback* conditions when a turn is linked to a previous turn (see Figure 5; GLMM (DV *complexity*) – interaction effect of condition by *continuesYN*; $F_{2,2657} = 21.818, p < 0.001$; significant pairwise interactions where *continuesYN=y* *interactive/playback* $t_{1,2657} = 4.410, p < 0.001$; *interactive/nominal pairs* $t_{1,2657} = 2.964, p = 0.003$). Developing each other’s suggestions leads to more complex ideas – but only if there is genuine interactivity.

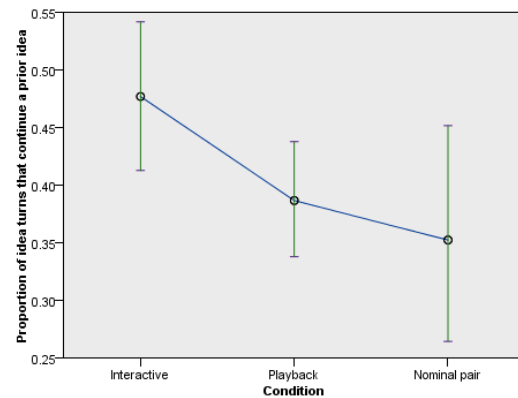


Figure 3: Marginal means of proportion of idea turns that continue a prior turn

Of the turns which continue another turn, there are also differences between the conditions in whether the idea that is continued is one that has been suggested within the same item (e.g. an idea for the object ‘brick’ is followed by another idea also for the object ‘brick’) or from a prior item (e.g. the idea ‘weapon’ may be suggested for ‘brick’ and then subsequently repeated or developed for ‘pencil’, or vice versa). *Interactive* dyads are more likely to be continuing an idea suggested within the same item, followed by those in the *playback* condition, with a lower probability of continu-

ing an idea within an item for the repetition or development of ideas in *nominal pairs* (see Table 3 and Figure 6; GLMM with `continuesSame` as DV; $F_{2,1124} = 8.880, p < 0.001$; pairwise effects: *interactive/playback* $t_{1,1124} = 2.800, p = 0.005$; *interactive/nominal pair* $t_{1,1124} = 3.470, p = 0.001$, *playback/nominal pair* $t_{1,1124} = 2.000, p = 0.046$). This means that while there is an effect of seeing extra locally relevant information in the *playback* condition, there is an additional effect in the *interactive* condition, where local context is key.

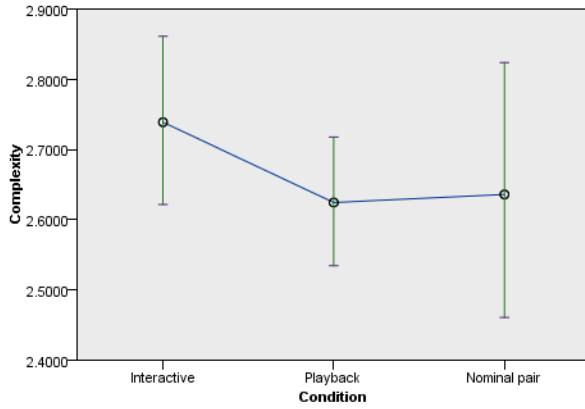


Figure 4: Marginal means of complexity of ideas

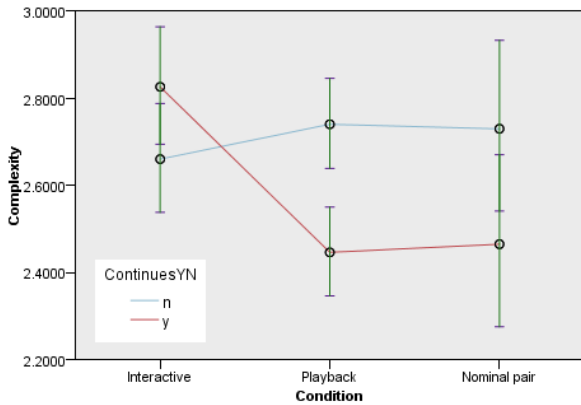


Figure 5: Marginal means of complexity of ideas by whether they continue a prior turn

Discussion

The results show that contrary to hypothesis 1, *playback* dyads do not produce significantly more or more complex ideas than non-interacting *individuals*. Similarly, contra hypothesis 2, although *interactive* dyads produce more turns in total, they do not produce more or more complex ideas than either *playback* dyads or *individuals*.

One explanation of this might conclude that the ‘extra’ turns produced by the *interactive* dyads mean that they are less efficient at generating ideas than those in the other conditions, given that all conditions generate the same number of uses. However, this would miss the qualitatively different

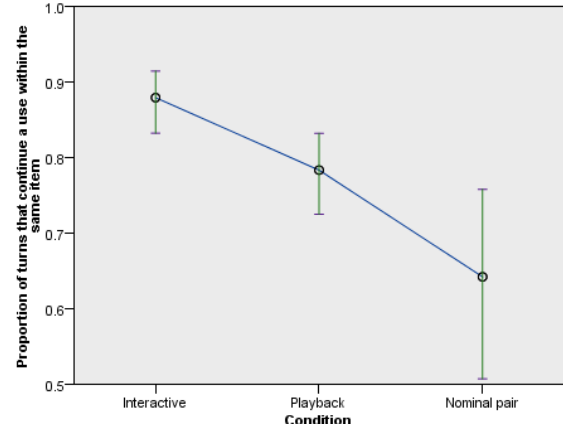


Figure 6: Marginal means of ideas turns which continue another turn by whether the turn is within item

idea generation process that conversation engenders, and ignore the fact that although participants in the *interactive* condition are doing more work (in terms of the total number of turns produced) this does not impact adversely on the number of ideas that they generate.

Additionally, these results may be influenced by the imposition of an arbitrary time limit. It is possible that *interactive* dyads would have continued to produce ideas after the five minutes was reached, whilst those in the *playback* and *individual* conditions may not have done. Future experiments could allow participants to indicate themselves when they felt they had exhausted their generation of ideas to test whether this was indeed the case.

In support of hypothesis 3, *interactive* dyads do build on each other’s contributions more than *playback* dyads, and the nature of their contributions are different, with turn sequence a crucial factor. This suggests that conversational mechanisms promote the creative exploration of a problem space. When people take advantage of the interaction to feed off one-another’s suggestions, they produce more complex solutions. This also suggests that a simple ‘idea’ count approach would miss the way that conversation can influence the process.

Interestingly, interactivity seems to change the nature of the exploration of the search space – allowing participants to probe deeper along a particular idea branch, but not necessarily to cover such a broad range of ideas. This indicates that interactivity may or may not be beneficial – depending on the specific goals of the problem. For the creative idea generation task reported here this type of deep search may be desirable, but for problem solving tasks with a correct answer this may be inappropriate. To investigate this, we are conducting experiments using the same methodology on a number of different types of task, such as the Remote Associates Task (e.g., find the word common to these three words: ‘show’, ‘life’, and ‘row’; answer ‘boat’), and the category fluency task (e.g.,

‘name all the animals you can think of’).

Of course, the preliminary results from this experiment do not show how specific conversational mechanisms could be shaping the form of specific ideas, however, anecdotally, participants in the *interactive* condition reported being more engaged in the task. A brief examination of the types of turns that participants make in the *interactive* condition suggests that this could be because some of the turns involve positive evaluative feedback (see e.g. (5) lines 3, 8 and 9), and several involve repetitions with modification (see e.g. (5) lines 3, 7, 11). Engagement could be further explored in future research, but this finding converges with large scale corpus studies that suggest that people build on each other’s turns to maintain the forward momentum of conversation (Healey, Purver, & Howes, 2014).

(5) Interactive dyad (Y & M); ‘barrel’

- 1 Y: a foot stool
- 2 Y: :P
- 3 M: a stool yes good one!
- 4 M: a coffee table
- 5 M: you could put glass on the top
- 6 Y: true
- 7 M: and make it a coffee table for a cottage
- 8 M: that would be quite nice :)
- 9 Y: that sounds like a nice idea
- 10 M: yeah i thought so :P
- 11 Y: we should make a business and sell them to cottage people.

In conclusion, the results point the way to a number of future research directions and indicate that it is not just exposure to someone else’s ideas that contributes to a qualitatively different and more complex form of problem solving, but the process of interaction itself.

Acknowledgments

This work was supported by an EPSRC grant EP/K503411/1.

References

- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.
- Clark, H. H. (1996). *Using language*. Cambridge University Press.
- Dennis, A. R., & Williams, M. L. (2003). Electronic brainstorming: Theory, research, and future directions. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity: Innovation through collaboration*. Oxford University Press.
- Dugosh, K. L., & Paulus, P. B. (2005). Cognitive and social comparison processes in brainstorming. *Journal of Experimental Social Psychology*, 41(3), 313–320.
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*.
- Gilhooly, K., Fioratou, E., Anthony, S., & Wynn, V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology*, 98(4), 611–625.
- Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York: Irvington Publishers.
- Healey, P. G. T., Purver, M., & Howes, C. (2014, June). Divergence in dialogue. *PLoS ONE*, 9(6), e98598. Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0098598> doi: 10.1371/journal.pone.0098598
- Healey, P. G. T., Purver, M., King, J., Ginzburg, J., & Mills, G. (2003, August). Experimenting with clarification in dialogue. In *Proceedings of the 25th annual meeting of the cognitive science society*. Boston, Massachusetts.
- Hill, G. W. (1982). Group versus individual performance: Are $N + 1$ heads better than one? *Psychological Bulletin*, 91(3), 517.
- Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment* (Vol. 53). John Wiley & Sons.
- Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annu. Rev. Psychol.*, 55, 623–655.
- Kohn, N. W., Paulus, P. B., & Choi, Y. (2011). Building on the ideas of others: An examination of the idea combination process. *Journal of Experimental Social Psychology*, 47(3), 554–561.
- Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and applied social psychology*, 12(1), 3–23.
- Nijstad, B. A., Stroebe, W., & Lodewijckx, H. F. (2002). Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of experimental social psychology*, 38(6), 535–544.
- Purver, M. (2001, October). *SCORE: A tool for searching the BNC* (Tech. Rep. No. TR-01-07). Department of Computer Science, King’s College London. Retrieved from <ftp://ftp.dcs.kcl.ac.uk/pub/tech-reports/tr01-07.ps.gz>
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Silvia, P. J. (2011). Subjective scoring of divergent thinking: Examining the reliability of unusual uses, instances, and consequences tasks. *Thinking Skills and Creativity*, 6(1), 24–30.
- Silvia, P. J., Martin, C., & Nusbaum, E. C. (2009). A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills and Creativity*, 4(2), 79–85.
- Ziegler, R., Diehl, M., & Zijlstra, G. (2000). Idea production in nominal and virtual groups: Does computer-mediated communication improve group brainstorming? *Group Processes & Intergroup Relations*, 3(2), 141–158.