

From Face-to-Face to Paired Oral Proficiency Interviews: The Nut is Yet to be Cracked

Parviz Birjandi

Faculty of Persian Literature and Foreign Languages, Islamic Azad University

South Allameh Street, Sa'adat Abad, Tehran, Iran

E-mail: pbirjand@yahoo.com

Marzieh Bagherkazemi (Corresponding author)

Ph. D. Candidate, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University

South Allameh Street, Sa'adat Abad, Tehran, Iran

E-mail: mbagherkazemi@gmail.com

Received: October 28, 2010 Accepted: November 10, 2010 doi:10.5539/elt.v4n2p169

Abstract

The pressing need for English oral communication skills in multifarious contexts today is compelling impetus behind the large number of studies done on oral proficiency interviewing. Moreover, given the recently articulated concerns with the fairness and social dimension of such interviews, parallel concerns have been raised as to how most fairly to assess the oral communication skills of examinees, and what factors contribute to more skilled performance. This article sketches theory and practice on two rather competing formats of oral proficiency interviewing: *face-to-face* and *paired*. In the first place, it reviews the related literature on the alleged disadvantages of the individual format. Then, the pros and cons of the paired format are enumerated. It is discussed that the paired format has indeed met some of the criticisms leveled at individual oral proficiency interviewing. However, exploitation of the paired format as an undisputable alternative to the face-to-face format begs the question.

Keywords: Contingency, Face-to-face/Individual OPI, Oral proficiency interview (OPI), Paired OPI

1. Introduction

Fairness in assessment in general and language assessment in particular can be of substantive concern to both assessors and those who are assessed once the vitality of immediate and far reaching consequences of assessment acts is brought to light. In other words, a fair assessment is tantamount to the assessor's staunch commitment to 'ethics' as an intrinsic part of his/her profession. Given this, a salient aspect of assessment is the validity and reliability of the procedures and measures utilized to assess individuals, and consequently the extent to which such procedures and measures make for the best performance of examinees. Nowhere in language testing and assessment have such issues been as much an eye sore as in the assessment of speaking. Heaton (1988) cogently has the point when he states that speaking "is an extremely difficult skill to test, as it is far too complex a skill to permit any reliable analysis to be made for the purpose of objective testing" (p. 88). Along the same lines, Foot (1999) mentions reliability, validity, being live and requiring the presence of an examiner, and also cost and time-efficiency considerations as four particularly problematic issues in foreign/second language speaking assessment.

Techniques to assess spoken language skills run the gamut from reading aloud through picture description to oral proficiency interviews (OPI) or face-to face interaction. Thanks to the communicative movement of the late 1970s, procedures at the OPI end of the continuum have enjoyed high status and intuitive appeal among speaking examiners and assessors. McNamara and Roever (2006) rightly observe that communicative language testing led to the predominance of 'face-to-face interaction' as the context of assessing spoken language skills. The 'communicative' tone has spelt many a definition of OPI.

Ross and Berwick (1992) define an oral proficiency interview as 'a sample of extended discourse, as a hybrid of interview and conversational interaction, and as an instance of communication across cultures' (p. 160). Central to this apparently simple definition are the mesmerizing CLT words 'interaction' and 'communication across cultures'. However, given the multi-faceted and complicated nature of these concepts as partially attested to by conversation

analysts, the long-standing controversy on how best, i.e. how most validly, reliably and fairly, to assess them is rarely unexpected.

Among different modes of OPI, the choice over the *individual* or *face-to-face* format- wherein an interviewer or examiner engages in an interview or conversation with an examinee- and the *paired* format- wherein two examinees orally interact with each other in the presence of two examiners - has long been a matter of hot debate. The paired format has followed and overtaken the individual mode and forms the sole or part of OPIs employed in most international tests of English language proficiency, including University of Cambridge ESOL examinations. The disposition running through the related literature is mostly toward the latter, while some studies have undertaken to bring into spotlight its downsides. The present study provides a general layout of both formats, and outlines several controversial issues haunting the paired format which have unquestionably passed the scrutiny of the oral proficiency assessment scholars.

2. Oral proficiency interviewing: A brief historical overview

Oral proficiency interviews have a fairly long history to them and “while objections have been (and continue to be) staged regarding numerous aspects of the Oral Proficiency Interview (OPI), there seems to be widespread agreement that it is the most appropriate tool for measuring oral proficiency” (Lazaraton, 1992, p.373) . However, Taylor and Wigglesworth (2009) concede:

Whether the interaction involves a test taker and test examiner/rater in the traditional individual format, or a pair or group of test takers, the co-constructed nature of the interaction, and the fact that co-participants’ contributions are inextricably linked, raises issues for language testers relating to construct definition, reliability and fairness (p. 328).

The developmental trend of OPI can be envisaged vis-à-vis that of language proficiency theorizing. The first oral proficiency interviews can be traced back to the late 1950s as evident in the development of the Foreign Service Institute’s “absolute proficiency scale” and its associated interview-based testing approach (Clark & Hooshmand, 1992) where the individual or face-to-face format was adopted as the norm. FSI’s scale was originally designed to evaluate the language proficiency of members of the US Foreign Service, and the FSI interview evaluated not only language proficiency, but also interpersonal and communication skills. Such interviews, according to Stanfield and Kenyon (1992) were grounded in the ‘psychometric-structuralist’ model of language proficiency as comprising the four seemingly distinct communicative skills of reading, writing, speaking and listening. This skills-based model was preoccupied with surface features of the language. Models of the sort dominated OPIs for a matter of 30 years and their indelible mark is manifest in the sustenance of the individual format well after the so-called ‘communicative revolution’ though moulded to meet its concerns.

However, with the advent of the communicative movement and in line with the push toward models of communicative competence from the late 1970s on, language testing, including oral proficiency assessment, underwent a breakthrough and was considerably influenced by increasingly broader conceptualizations of communication. Such ‘communicative-sociolinguistic’ models came to appreciate pragmatic, strategic and contextual aspects of proficiency and gave rise to a lot of issues surrounding oral proficiency interviewing previously unattended to. Although the individual mode never left the stage, along with emphasis on pair work in language learning contexts came a growing interest in paired language assessments, particularly in the context of oral proficiency interviewing (Taylor & Wigglesworth, 2009).

3. The individual or face-to-face format

As mentioned earlier, the individual format is characterized by the presence of an examinee and an interlocutor who mostly also acts as the rater/examiner. To exemplify the individual format, one can refer to the OPI of the American Council on the Teaching of Foreign Languages (ACTFL) which has been in use and under much criticism since 1982. Having as its antecedents the FSI (Foreign Service Institute) and ILR (Interagency Language Roundtable) OPIs, the ACTFL with its continually revised guidelines emphasizes “authentic language use in communicative contexts” (Henning, 1992, p.365). This five-level interview has a highly structured nature that makes it essentially different from the paired format (Yoffee, 1997):

- (1). *Warm-up*, to make the interviewee feel comfortable and to familiarize him/her with the setting;
- (2). *Level check*, to unearth his/her ability to manipulate tasks at a particular level;
- (3). *Probes*, to elicit responses at a higher level, and to reveal weaknesses;
- (4). *Role play*, to confirm the testee’s level;
- (5). *Wind-down*, to come down to a level suiting the testee and end the interview on a positive note;

Several criticisms have been leveled at the ACTFL and other similar individual formats of OPI, among which the following stand out:

1). Asymmetry: ACTFL and other similar individual formats of OPI are *asymmetrical* in that they exert power over the interviewee in terms of question formation, discourse trajectory, choice of content, and 'moves' distribution across the interviewer and the interviewee. In this regard, van Lier (1989) states: "In a sense, in asymmetrical discourse, miscommunication and pragmatic failure are *by definition* the controlling party's responsibility" (p. 499), and this mode of OPI has the interviewer as the controlling party. Accordingly, it has been pointed out that the individual mode is not *conversational* in nature and can only measure with accuracy interview proficiency or 'performance in context' and not general oral proficiency or conversation skills. In a similar vein, McNamara & Roever (2006) state that face-to-face interaction takes place within an 'interaction order', i.e. a socially and culturally, not necessarily linguistically, regulated face-to-face domain, and the status superiority of the interviewer has a determining influence on the performance of the interviewee.

2). Pseudo-contingency: The individual format of OPI has also been criticized for creating false contexts as in role-plays and therefore being pseudo-contingent. Although this problem can also be raised with the paired format, the asymmetrical nature of the individual format exacerbates the issue. The uneven distribution of power characteristic of such an OPI mode forestalls reactive and mutual contingencies which mark real conversations. Figure 1 is an illustration of four classes of social interaction in terms of contingency, adopted and adapted from Jones and Gerard (1967, p. 507):

- *asymmetrical contingency*, which describes the type of interaction found in traditional teaching and interviewing;
- *pseudo-contingency*, which describes speech events such as role plays and rituals (e.g. greetings);
- *reactive contingency*, as is the case with rambling conversations;
- *mutual contingency*, which typifies negotiations, serious discussions, etc.

It is evident from the representation that the paired format where the problem of asymmetry and imposition can be dealt with has greater potential in inducing reactive and, more significantly, mutual contingencies.

3). Negative washback on classroom practices: It is generally believed that "oral tests can have an excellent backwash effect on the teaching that takes place prior to the tests" (Heaton, 1988, p. 89). Yoffee (1997) states "...the washback effect [of oral tests] on classroom teaching has been positive as the practitioners place more emphasis on speaking, encouraging student oral production in class" (p. 10). However, the positive washback of the individual format has been called into question on the grounds that it sustains unequal power distribution and imposition in the classroom with teachers being the main initiators and students mostly only responding and receiving feedback on their responses (e.g., Lantolf & Frawley, 1988).

4. The paired format

Thanks to the findings of conversation analysts and owing to the growing awareness of the issues outlined earlier in this article, individualistic theories of language proficiency have been challenged by social views of performance, which maintain, in essence, that coherence, meanings, identities and events are co-constructed by interlocutors, and that the context of an interview is influenced by the presence of an interlocutor. Such views also take issue with unequal move opportunities present to interlocutors and candidates, and the influence of the interviewer's idiosyncratic accent, speech style, personality, functional pitch, questioning and feedback provision techniques and also topical focus on examinees' performance. The '*joint construction of performance*' can be said to amount to the influence of interlocutors on discourse outcomes and assessment results as the main source of variation (McNamara and Roever, 2006). In this regard, McNamara asserts that "the age, sex, educational level, proficiency or native speaker status and personal qualities of the interlocutor relative to the same qualities in the candidate are all likely to be significant in influencing the candidate's performance" (1996, p. 86).

In consequence, performance on conversational tasks cannot be directly inferred from performance on individual oral proficiency interviews, and the validity of inferences must be established by demonstrating the common features of the two performance situations and by providing empirical evidence. One of the solutions offered to address this issue is the paired or group mode of oral proficiency interviewing wherein two (or more) examinees engage in an oral interaction with each other in the presence of two examiners, one acting as an assessor and the other as an interlocutor. In other words, the paired format is marked by peer-peer interaction rather than or as well as examiner-examinee interaction (Taylor and Wigglesworth, 2009). However, the interlocutor has a more limited role compared with that in the individual format:

Typically the interlocutor explains the tasks to the candidates, engages them in conversation during the introductory stage of the test, asks them to explain their solution to any joint task, and acts as time-keeper. The assessor listens to the candidates and assesses them on the evidence of their performance in the tasks, against the established criteria. Assessors may, towards the end of the test, talk to the candidates for the purpose of 'fine tuning' the assessment (Foot, 1999, p.39).

4.1 Advantages of the paired OPI format

Upon browsing the related literature, the controversial rationale behind the rapid takeover of the paired mode can be summarized as follows:

(1). It is "psychologically easier" for both examinees and examiners. It reduces the pressure on an individual examiner who also acts as the interlocutor as is the case with the one-to-one format. As for examinees, familiarity helps share the anxiety. Even when they do not know each other, the information gap induced resembles that in real-life conversations (Heaton, 1988; Wallis, 1995). Therefore, it is no surprise that some studies have substantiated the claim that students like pairings (Egyud & Glover, 2001).

(2). Individual examiner bias is compromised, and marker reliability enhanced (Foot, 1999). One of the particularly problematic issues regarding oral proficiency interviewing in general is interviewer/examiner reliability; Calderbank and Awwad (1988) state that the reliability of an OPI, be it individual or paired, can be enhanced through rigorous interviewer training and the development of viable assessment instruments based on communicative criteria. The paired format is presumably advantaged in this regard owing to the presence of two examiners whose ratings can be pooled or averaged to obtain a compromise.

(3). It elicits a more varied pattern of conversation; there are three patterns of interaction in the paired format namely candidate-candidate, candidate-interlocutor and candidates-interlocutor. Accordingly, it is generally stated that such patterns, with the resultant greater range of speech events, allow the candidates to show their best. This is not the case with the individual mode where interviewer interventions can sometimes be debilitating rather than facilitative since the examinee performs on a different level from the interviewer who exerts more control (Foot, 1999; Egyud & Glover, 2001). In her study, Brooks (2009) came up with the following conclusion:

When test-takers interacted with other students in the paired test, the interaction was much more complex and revealed the co-construction of a more linguistically demanding performance than did the interaction between examiners and students. The paired testing format resulted in more interaction, negotiation of meaning, consideration of the interlocutor and more complex output (p. 341).

(4). Pairing helps to produce better English than one-to-one format. The latter is more like an interrogation in which inequality of partners is more outstanding leading to a limited range of speech acts and artificiality; in the individual format, initiation is exclusive to the interviewer and, unlike the paired format, is limited to the interlocutor-interviewer interaction pattern (Egyud & Glover, 2001).

(5). The paired format is more likely to induce positive washback on classroom practices and support good teaching since it encourages pair and group work, and reflects realistic student-student interaction (Egyud & Glover, 2001).

4.2 Issues in the paired OPI format

The paired mode has been in use since the 1980s, and the fact that it is now part of four UCLES (University of Cambridge Local Examinations Syndicate) exams, namely PET (Preliminary English Test), KET (Key English Test), FCE (First Certificate in English), and CAE (Certificate in Advanced English), attests to its widespread take-up as a safe and sound surrogate for the individual format. Toward the end of the 90s, the controversy arose as to whether the wider use of the paired format was justified. Foot (1999) regrets "... the lack of published research evidence, and of results from the monitoring of these tests to support their introduction and wider use" (p. 36). Several controversial issues surround the paired mode of oral proficiency interviewing. Some have been briefly pointed out in the literature (e.g., Foot, 1999; Fulcher, 2003; Norton, 2005). This section of the paper provides a summary overview of issues raised against the paired OPI format which render some of its presumed advantages at least worthy of closer scrutiny. One point worth mentioning is that the arguments presented are posed as unresolved points of contention and need to be substantiated by empirical research:

1). Does 'easier' mean 'better performance'? While some studies suggest that the paired mode results in higher scores than the individual mode, there is no reason to think that it is 'the relaxed atmosphere' of a paired oral proficiency interview that induces better performance on the grounds that the relationship between stress and performance is too complex to be sketched as a straight causal link. Moreover, while some proponents of the paired mode have brought up 'anxiety sharing' and consequently 'anxiety reduction' as its support, it can be equally stated

that anxiety is generally ‘contagious’, a presumption that is intuitively more appealing (Foot, 1999) Accordingly, even the assumption of an anxiety-free and a more relaxed interview ambience is at best questionable.

2). Should the candidates know each other? ‘Candidates’ familiarity’ has proven to raise scores on oral proficiency tests (Ildikó, 2002). Norton (2005) believes familiarity allays anxiety, enhances fluency and interactive communication, leads to better task achievement, equal participation, and more talk. However, knowing or not knowing the other candidate leads to two different kinds of tests and the problem is how to strike a balance between the *relaxation* which familiarity induces and the *information gap* characteristic of real-life conversations which unfamiliarity results in.

3). Should candidates share the same first language? If they don’t, problems of comprehensibility and trying to get tuned to the other candidate’s pronunciation and syntax are unavoidable. Therefore, where both types of pairs with the same and different first languages are possible, the same first language is likely to raise scores. In this regard, Lazaraton (1991) states that transfer of one’s L1 habits in terms of turn taking and topic initiation expectations on the interview tends to compromise performance. Taking a more panoramic view, sociocultural and pragmatic competencies which are important determiners of successful performance can be to a large extent influenced by one’s first language norms.

4). Should the candidates be of the same level of proficiency? This is a circular problem, since to assess speaking we need candidates who are of a comparable speaking proficiency level. A related concept is ‘appropriation’ meaning that candidates appropriate syntactic structures and lexical items from each other’s discourse. Accordingly, less proficient candidates, when paired with higher level candidates, may be at an advantage (Norton, 2005). In a related study, Iwashita (1996) found an interlocutor effect in terms of the proficiency level of paired candidates on the discourse produced but not on the scores assigned.

5). What should the nature of the social relationship between candidates, in terms of age, gender, social class and profession be? Do differences in such respects influence performance? The existing literature on paired testing tends to resist the argument that they do, but appealing to everyday experience, Foot (1999) argues that they definitely have an effect. As an example, Norton (2005) asserts that examinees of the same gender generally show a more equally distributed contribution.

6). Should candidates in a pair be matched on their personality traits? This is where the one-to-one format seems to be more advantageous. Heaton (1988) states that the paired format gains in validity if candidates with similar personality traits are paired with each other. However, practicality concerns associated with such an undertaking cannot be too greatly emphasized. Foot (1999) asks if one candidate is reserved and the other domineering, how is such information reflected in the final assessment? Van Lier (1989) points to this OPI validity threat when he bewares assessors of mistaking a reserved or ‘will-not-talk’ candidate for a ‘cannot-talk’ candidate.

7). Are candidates’ hidden intentions to help or fail a friend taken into account? It is generally admitted that “the co-participants each contribute to the interaction and so their performances are inextricably linked” (Brooks, 2009, p. 342). While Egyud and Glover (2001) deem ‘cooperation’ between candidates in paired testing as ancillary, Foot (1999) believes this cooperation might lead more able candidates to intentionally ‘underperform’, i.e. to tune in their performance to their partners’ out of sympathy, or apply ‘partner-failing’ conversation strategies.

8). To what extent is the examiners supposed to intervene and to encourage candidates to seek clarifications in the case of incomprehensible or uncomprehending candidates? How is such information reflected in the final assessment of the disadvantaged candidate? (Foot, 1999). One can argue that detailed guidelines can be set out and followed uniformly by all examiners to resolve the issue. However, each interview is a unique interaction situation the details of which cannot be fully predicted in advance. Accordingly, although one can postulate general guidelines, the idiosyncratic nature of each instance of OPI precludes prescribing examiners with a fixed set of ‘how-to’s. A related problem is that taking a back stance, assessors pass on to the learners the feeling that the interaction event is in fact artificial.

9). How can one ensure inter-marker reliability? One being a participant, the other only a spectator, the two examiners might disagree in their assessments even if they apply the criteria uniformly despite postulations as to the greater marker reliability of the paired format compared to the singleton mode.

10). Do the interaction patterns of the paired mode result in a wider range of speech events? Speech events (argument, description, discussion, narrative, and opinion) can be induced by any pattern of conversation. It goes without saying that because test-takers generally do not have any training in conducting oral proficiency interviews (Luoma, 2004), they may have difficulty managing the interaction. Unless they are well-matched, the candidates,

particularly when they are inexpert, cannot sustain a discussion; and without examiners' interventions, samples of performance will be inadequate for the purpose of assessment.

5. Conclusion

Outlining consensus and controversy over the use of individual and paired modes of oral proficiency interviewing, the present article chimes with calls for 'critical language testing', i.e. taking a critical stance toward language assessment procedures and measures. Oblivious acceptance and adoption of such measures and procedures at face value and just because of their widespread take-up after discrediting those already in use amounts to 'ignorant professionalism' in the era of critical assessment. Needless to say, such measures need to be screened for possible sources of bias and 'unfairness'.

Given the fact that "today, OPIs are used by academic institutions, government agencies, and private corporations for many purposes: academic placement, student assessment, program evaluation, professional certification, hiring, and promotional qualification" (Swender, 2008, p.520), the stakes involved in them are very high. The list of issues discussed in terms of the paired mode of OPI is not exhaustive and upon contemplating such a procedure for assessing spoken language skills several others surface. The paired format might indeed be more beneficial than the individual format, but before such a claim can be made several issues inundating the paired mode should be resolved through empirical research. Researchers are called upon to carry out empirical studies comparing the individual and paired OPI formats in terms of the questions posed and provide empirical evidence on the latter's presumed priority.

References

- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26 (3): 341-366.
- Calderbank, M. & Awwad, J. (1988). Testing oral communication: Why and how. *System*, 16(1):49-59.
- Clark, J.L.D. & Hooshmand, D. (1992). Screen-to screen testing: An exploratory study of oral proficiency interviewing using video conferencing. *System*, 20 (3): 293-304.
- Egyud, G. & Glover, P. (2001). Oral testing in pairs: A secondary school perspective. *ELT Journal*, 55 (1):70-76.
- Foot, M. (1999). Relaxing in pairs. *ELT Journal*, 53(1): 36-41.
- Fulcher, G. (2003). *Testing second language speaking*. Harlow: Pearson.
- Heaton, J.B. (1988). *Writing English language tests*. New York: Longman Inc.
- Henning, G. (1992). The ACTFL oral proficiency interview: Validity evidence. *System*, 20 (3): 365-372.
- Ildikó, C. (2002). Is testing speaking in pairs disadvantageous for students? A quantitative study of partner effects on oral test scores. *novELTy*, 9 (1): 22-45.
- Iwashita, N. 1996. The validity of the paired interview in oral performance assessment. *Melbourne Papers in Language Testing*, 5(2): 51-65.
- Jones, E.E. & Gerard, H.B. (1967). *Foundations of social psychology*. New York: John Wiley & Sons.
- Lantolf, J. P. & Frawley, W. (1988). Proficiency: Understanding the construct. *Studies in Second Language Acquisition*, 10(2), 181-195.
- Lazaraton A. 1991. *A conversation analysis of structure and interaction in the language interview*. Unpublished PhD Dissertation, University of California.
- Lazaraton, A. (1992). The structural organization of a language interview: A conversation analytic perspective. *System*, 20 (3): 373-386.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- McNamara, T. (1996). *Measuring second language performance*. Harlow: Longman Inc.
- McNamara, T. & Roever, C. (2006). *Language testing: The social dimension*. USA: Blackwell Publishing.
- Norton, J. (2005). The paired format in the Cambridge Speaking Tests. *ELT Journal*, 59 (4): 287- 297.
- Ross, S. and R. Berwick. 1992. 'The discourse of accommodation in oral proficiency interviews'. *Studies in Second Language Acquisition*, 14(2): 159-76.
- Stanfield, C.W. & Kenyon, D.M. (1992). Research on the comparability of the oral proficiency interview and the simulated oral proficiency interview. *System*, 20 (3): 347-364.

Swender, E. (2008). Oral proficiency testing in the real world: Answers to frequently asked questions. *Foreign Language Annals*, 36 (4): 520-526.

Taylor, L. & Wigglesworth, G. (2009). Are two heads better than one? Pair work in L2 assessment contexts. *Language Testing*, 26 (3): 325-339.

van Lier, L. 1989. 'Reeling, writhing, drawing, stretching and fainting in coils: Oral proficiency interviews as conversation'. *TESOL Quarterly*, 23(3): 489-508.

Wallis, D. 1995. 'Testing in twos—the paired format in oral examinations'. *UCLES presentation, IATEFL Annual Conference*, New York.

Yoffee, L. (1997). An overview of ACTFL proficiency interviews: A test of speaking ability. *JALT Testing & Evaluation SIG September Newsletter*. [online] Available: http://www.jalt.org/test/yof_1.htm.

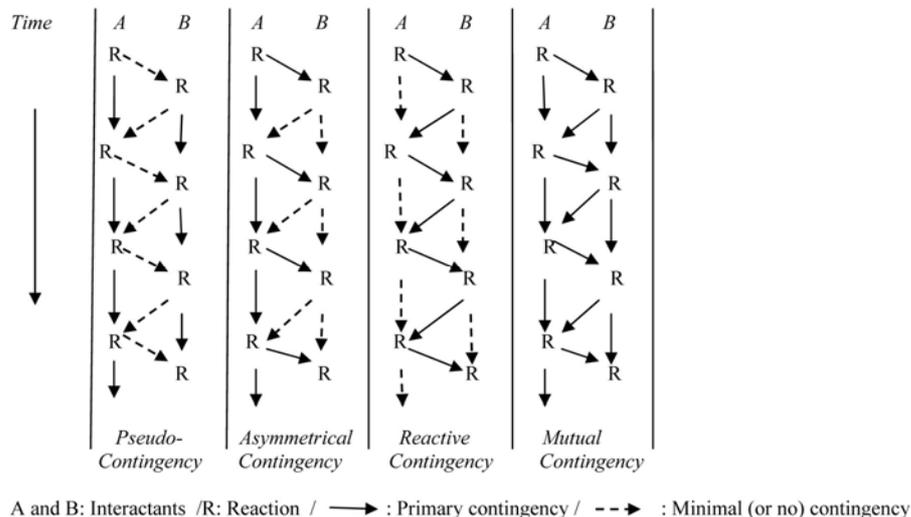


Figure 1. Classes of social interaction in terms of contingency