We used West and Kenny’s (2011) Truth-and-Bias (T&B) model to examine how accurately group therapists judge their group members’ alliances, and the effects of therapist-patient congruence in alliance ratings on patient outcomes. Were considered: (a) directional bias – therapists’ tendency to over- or underrate their clients’ alliances, (b) truth strength – clients’ alliance ratings, and (c) bias strength – therapists’ tendency to conflate their alliance ratings for a specific group member with the average alliance ratings for the other members of the group. There were 118 obese adult patients with binge-eating disorder that were treated by 8 therapists with Emotionally Focused Group Therapy. Outcomes were operationalized as pre- to postchanges in: health-related quality of life, binge eating, and psychological distress. Patients’ and therapists’ working alliance were assessed after the 2nd, 10th, and last (20th) group therapy sessions. (a) There was no significant congruence between group therapists’ and members’ ratings of alliance; (b) therapists’ ratings of an individual group member’s alliance were significantly related to therapists’ ratings of the other group members’ alliance in early sessions but unrelated in later sessions; and (c) the relationship between therapists’ alliance ratings and bias strength was weaker when patient binge eating outcomes improved. Group therapists adopted a “better safe than sorry” strategy by underestimating the strength of their group members’ alliances. Therapists had a tendency to judge each group member’s individual alliance based on the aggregated alliance of the other group members. Improvement in patient binge eating outcomes was related to therapists overcoming this tendency.

Keywords: truth and bias model, emotionally focused group therapy, binge-eating disorder, working alliance, therapist accuracy

Supplemental materials: http://dx.doi.org/10.1037/pst0000042.supp

The therapeutic alliance is one of the most researched concepts in individual (Horvath et al., 2011) and group (Burlingame, Strauss, & Joyce, 2013) psychotherapy. Therapeutic alliance as well as several group relationship constructs in group psychotherapy is comprised of three structural components, including the relationship between the therapist and an individual member, an individual member and other group members, and an individual member and the group as a whole (Burlingame, McClendon, & Alonso, 2011). Research has demonstrated that group alliance is related to therapeutic outcome in patients with different psychiatric disorders, such as borderline personality disorder (Marziali, Munroe-Blum, & Mc Cleary, 1999), anxiety and depression (Budman et al., 1989), complicated grief (Joyce, Piper, & Ogrodniczuk, 2007), and binge eating disorder (Tasca, Balfour, Ritchie, & Bisada, 2007). Given the structural complexity of group settings, research on group alliance still needs to address the interplay among these structural elements (i.e., therapist-members-group-as-a-whole). Only a few studies focused on the member-leader and member other-group-members dimensions (Burlingame, Fuji rman, & Johnson, 2002). Recent studies evidenced that patients and therapists represent a strong source of variation accounting for patient rated alliance (Bakali et al., 2010) and that patient symptom reduction was greater when there was agreement between a group member and other group members that their alliance to the group was strong (Lo Coco, Gullo, & Kivlighan, 2012).

Commonly, the alliance is defined as having three key elements that clients and therapists share: an interpersonal and emotional bond, an agreement on the tasks of therapy, and an agreement on the goals of therapy (Bordin, 1979). Implicit in this definition is
that there is congruence between therapists and clients on the tasks and goals of therapy, and that this congruence has an impact on their bond. That is, therapists, individuals, and groups must be on the same page in terms of how therapy will work and what to expect. This affects how much they feel understood by the other, and ultimately will affect whether the client benefits.

Mary’s group therapist believes that she and Mary have a strong working alliance. The therapist asks Mary to give some difficult feedback to Brian, another group member, about Brian’s dismissing interpersonal style. However, Mary feels distant from her group therapist and does not agree that providing feedback to others is helpful. That is, Mary and her group therapist do not have an agreement on the useful and important tasks of group therapy. Mary refuses her group therapist’s request and subsequently feels even more distant from and misunderstood by her therapist. This reduces the interpersonal and affective bond between the group therapist and Mary. If Mary’s group therapist continues to view their working alliance differently than does Mary, and so continues to use interventions that do not fit for Mary, their relationship will suffer and Mary likely will not benefit from the group.

In this study we assume that: (a) group therapists make judgments about the collaborative relationship with individual group members, and (b) therapists choose how to intervene with a group member based in part on judgments about the quality of their alliance. Further, if the therapist’s judgment about their alliance aligns with the group member’s judgment (i.e., they agree on tasks and goals of therapy and they share a bond), then the therapist can intervene in ways that promotes the group member’s work. If the therapist’s judgment about the alliance is not congruent with the group member’s judgment, then the therapist’s interventions will be inappropriate, which may alienate the group member and reduce the chances for a good outcome.

This study draws data from an effectiveness trial of Emotionally Focused Group Therapy (EFGT) for binge-eating disorder (BED; Compare & Molinari, 2012; Compare & Tasca, 2014). EFGT focuses primarily on the client’s relationship with and regulation of difficult emotions to help the client to manage and overcome the fear associated with social situations. EFGT also focuses on the relationship between client and therapist. In an effectiveness trial, Compare and colleagues (2013) demonstrated that EFGT resulted in reduced binge eating among obese individuals with BED. BED is the most common of the eating disorders (i.e., 3.5% population lifetime prevalence), is comorbid with obesity and lower health related quality of life, and is associated with high rates of depression, anxiety, and interpersonal problems (Tasca et al., 2012). Binge eating in those with BED can result from difficulty regulating negative emotions, which in turn may be the result of interpersonal problems (Ivanova et al., 2015). Therefore, EFGT may be effective because it targets both emotion regulation and interpersonal distress. In a study of a different interpersonally based group therapy for BED, members of our team demonstrated that an increasing therapeutic alliance across group sessions was associated with reduced binge eating and depressive symptoms at post treatment (Tasca et al., 2007). Although this study was suggestive of the important role of therapeutic alliance in the group treatment of BED, there is currently no study to assess the way in which alliance is experienced by these clients’ group therapists, the congruence with which the alliance is experienced, and the impact of congruence on client outcomes.

In individual treatments, early studies examining the theoretical predictions about working alliance congruence obtained mixed results (e.g., Kivlighan, 2007). However, recent, and more methodologically sophisticated studies showed that agreement between therapists and clients in their alliance ratings was related to both counseling and session outcome (Marmarosh & Kivlighan, 2012). Only two studies have examined congruence between group therapists’ and group members’ views of the group relationship and reported inconclusive findings. Jenkins et al. (1971) found no relationship at any point in time between therapist and members on ratings of the therapeutic relationship and therapist effectiveness in two groups. However, congruence was stronger for ratings of therapist use of techniques and direction. Chapman et al. (2012) compared group leaders’ and group members’ perceptions of positive bonding relationships, positive working relationships, and negative relationships. Chapman et al. (2012) concluded that therapists were not accurate in their perceptions of how group members perceived group therapeutic relationships, especially in groups with seriously mentally ill patients. However, the authors found that group therapists working at a university counseling center were more accurate in later rather than earlier group sessions. It is noteworthy that neither Jenkins et al. (1971) nor Chapman et al. (2012) linked congruence in group leader-group member relational perceptions to client outcome as was done in an individual counseling study by Marmarosh and Kivlighan (2012). Moreover, Jenkins et al. (1971) and Chapman et al. (2012) only reported correlations between therapists and clients at each separate time point; that is, they did not use growth curve modeling to examine how therapist-group member congruence changed across time. As we describe below, Atzil-Slonim et al. (2015) did use growth curve modeling to examine within-person examination of how therapist-client congruence in alliance ratings during individual therapy varies over time.

There appears to be stronger support for the occurrence of congruence in relational perceptions in the individual therapy literature than in the group therapy literature. This may occur because congruence in relational perceptions may be harder to achieve in group therapy because therapists have to track multiple relationships. As indicated, more recent advances in statistical modeling of longitudinal and grouped data allow researchers to address these complex design issues (Atzil-Slonim et al., 2015; Marmarosh & Kivlighan, 2012). Hence, in this study we examined: (a) if there was significant temporal congruence in the perception of the working alliance between group therapists and group members, and (b) if temporal congruence in working alliance ratings between group therapists and each of their group members was related to the group members’ treatment outcomes.

**The Truth and Bias Model**

It is the therapist’s responsibility to understand their client’s experience of the alliance (Atzil-Slonim et al., 2015). Hence working alliance agreement can be formulated as therapists’ attempts to accurately perceive his/her clients’ rating of the working alliance. Like Atzil-Slonim et al.’s (2015) study, we used the “Truth-and-Bias” (T&B) model (West & Kenny, 2011) to examine therapists’ ability to accurately perceive their clients’ alliance. In T&B terminology we treated group members’ ratings of the working alliance with their group therapist as the “truth,” and the group
therapists’ ratings of the working alliance with each of their group members as “judgments” about the group member’s working alliance.

Working alliance congruence can be operationalized in terms of: (a) the mean difference in group members’ and group therapists’ working alliance ratings, or (b) the correlation between these ratings. An important advantage of the T&B model is that it simultaneously examines both the mean difference in, and the correlation between the working alliance ratings of group therapists and group members.

A generalized version of the T&B model, originally described by West and Kenny (2011) can be formally expressed for perceiver (i.e., group therapist) “i” as:

\[ J_{Ci} = b_0 + tC_{Ti} + bB_{Ci} + E_i, \]  

This generalized version of the T&B model is different from the T&B model examined in the current study because the generalized version does not address the different levels of data nesting (i.e., sessions within members, and members within groups). We use this generalized model to illustrate the relationships examined in the T&B model. The specific T&B model used in this study is presented supplementary online material.

In the model shown above, \( J_{Ci} \) is a group therapist’s judgment of one group member’s alliance; \( b_0 \) is the directional bias (i.e., the degree to which the group therapist is biased to perceive the group member as having a stronger or weaker alliance than the group member reports); \( t_{C_{Ti}} \) is the truth variable (i.e., the group member’s alliance rating); \( t \) is the truth force (i.e., the strength of the effect of the group member’s actual alliance rating on the group therapist’s judgment of the group member); \( B_{Ci} \) is the bias variable (in our case, the aggregated alliance ratings for the other group members, excluding the focal group member’s alliance rating); \( b \) is the bias force (i.e., the strength of the effect of the other group members’ alliance ratings); and \( E_i \) is random error. The “c” subscripts indicate that the each of the variables is centered using the grand mean of the truth variable.

In a T&B study of individual therapy, Atzil-Slonim et al. (2015) found significant negative directional bias; that is, the therapists’ ratings of the working alliance were significantly lower than their clients’ ratings of the working alliance. Atzil-Slonim et al. (2015) theorized that the negative directional bias resulted from therapists adopting a “better safe than sorry” approach. That is, therapists may see that it is better for the relationship if they under-rather than over-rate the alliance. Overrating the strength of the alliance may result in the therapist asking the client to do things (e.g., disclose, take risks) that the client is not ready to do (Marmarosh & Kivlighan, 2012). This was illustrated in the example we presented above of Mary and her group therapist. Based on Atzil-Slonim et al. (2015) better safe than sorry theory we hypothesized that there will be: A directional bias effect in which EFGT therapists’ working alliance ratings for a specific group member will be significantly lower than that group member’s working alliance ratings.

Therapist and group members have different therapy experiences and different roles, and so therapists and clients may interpret their alliances differently. Both Atzil-Slonim et al. (2015) and Kivlighan and Shaughnessy (1995) found a significant truth effect (i.e., clients’ working alliance ratings were positively related to their therapists’ working alliance ratings). However, neither Jenkins et al. (1971) nor Chapman et al. (2012) found a truth effect for alliance ratings in group therapy. Given these contradictory findings we examined the truth effect without making a hypothesis.

As noted above, group therapists have multiple alliances to track. Hamilton and Sherman (1996) showed that information related to individuals and information about aggregates (groups of individuals) is stored in separate locations in memory. The cognitive demand of tracking all of these multiple alliances, coupled with this information being stored in different locations, can result in resource depletion (Baumeister, Bratslavsky, Muraven, & Tice, 1998). One way that group therapists may reduce resource depletion is to have a heuristic for judging their working alliances with individual group members. Specifically, we argue that group therapists develop a general sense of their ability to form working alliances with the members of their group, which may bias their ability to accurately judge their working alliance with a specific group member. Using Kenny, Mannetti, Pierro, Livi, and Kashy (2002) actor partner model for small groups, we define bias as the aggregated alliance ratings of the other group members excluding the focal member. Therefore, we hypothesized that: there will be a significant positive relationship between the EFGT therapist’s ratings of the working alliance with a specific member and the therapist’s aggregated ratings of the alliances with the other group members.

Group Member Outcome as a Moderator of Directional Discrepancy and of Truth and Bias Strength

In the T&B model variables that can affect directional discrepancy and truth and bias strength are examined as person-level moderators of the T&B effects. In this study we used a moderator analysis to examine the relationship between group member outcomes in EFGT and directional discrepancy and truth and bias strength. As discussed at the outset, we conceptualize congruence as an agreement on tasks and goals of therapy, which in turn will have an impact on the bond between client and group therapist. We also based our hypotheses on Pepinsky and Karst’s (1964) theory that implies that group members will have better outcomes when group therapists accurately judge the state of their group members’ individual alliances. Pepinsky and Karst’s (1964) theory is at odds with Atzil-Slonim et al.’s (2015) better safe than sorry model. Pepinsky and Karst’s theory predicts that accuracy is related to better outcome whereas Atzil-Slonim et al.’s model predicts that a particular type of inaccuracy, that is, therapists’ underrating of client alliance, is related to better outcome. We contend that Pepinsky and Karst’s (1964) theory addresses correlational accuracy whereas Atzil-Slonim et al.’s (2015) better safe than sorry model is addressing mean difference accuracy. As noted above in a study of individual treatment, Marmarosh and Kivlighan (2012) found that therapist correlational accuracy was related to client outcome. In the current study, correlational accuracy of judgments is operationalized as greater truth strength and lesser bias strength. Therefore, we hypothesized that: (a) there will be a stronger truth effect for clients with BED who improved and a weaker truth effect for clients who did not improve; and (b) there will be a weaker bias effect for clients with BED who improved and a stronger bias effect for clients who did not improve.
Research suggests that therapists’ underrating of the alliance relative to clients is related to a stronger truth effect and better session- and treatment-outcome (Atzil-Slonim et al., 2015). Therefore, we hypothesized that: group member symptom improvement will be negatively related to directional bias.

In the current study of EFGT for BED, group therapist and group member working alliance ratings were measured across time. In addition, our bias variable, which was an aggregation of the other group members’ working alliance ratings, also consisted of ratings across time. Because positive (or negative) truth effect or bias effects could result from both group therapists and group members increasing their working alliance over time, it was important to examine the effects of time by including: (a) a main effect for time, (b) a time by truth interaction, and (c) a time by bias interaction in our models.

The time by truth and time by bias interactions also have important substantive meanings. With increased interpersonal and emotional awareness resulting from EFGT, others group members’ alliance to the therapist should be easier to judge as time goes on. In addition, the emphasis on differentiation and individuation in the latter stages of EFGT should make it easier for the group therapists to recognize the distinctive alliances that they have with each of their group members. Based on these conceptualizations of EFGT, we hypothesized that: (a) the association between the group therapists’ alliance ratings for individual group members and the truth variable (the member’s actual alliance rating) will increase over time; and (b) the association between the group therapists’ alliance ratings for individual group members and the bias variable (the aggregated alliance ratings of the other group members) will decrease over time.

Method

Participants

The study involved obese adult patients (N = 118) meeting Diagnostic and Statistical Manual for Mental Disorders-Fourth Edition (DSM–IV) research criteria for BED (American Psychiatric Association, 1994). Participants were referred to an outpatient department at San Michele Hospital in Milan, Italy, for the treatment of eating disorders by family doctors or other clinicians between 2008 and 2012. To be eligible for the study, patients were required to: (a) be between 35 and 60 years old, (b) have a body mass index (BMI) of 30 or greater, and (3) meet DSM–IV criteria for BED. Patients attended EFGT or Combined Therapy (CT) of EFGT plus Dietary Counseling for 5 months. Of the 126 patients recruited for the study, 8 discontinued the group treatment. The 118 participants (56 male and 62 female) had a mean age of 50.9 years (SD = 4.91) and a mean BMI of 33.03 (SD = 2.01).

Procedure

The general flow of participants through the study and pretreatment characteristics was described in previous reports (Compare, Calugi, Marchesini, Molinari, et al., 2013; Compare, Calugi, Marchesini, Shonin, et al., 2013; Compare & Tassca, 2014). Diagnostic assessments of BED were completed by trained clinical psychologists. Participants who met inclusion criteria provided written informed consent and were assigned to one of the treatment programs. Treatment condition assignment was conducted by consensus among a team of clinicians and was based on psychometric test results (described in Compare, Calugi, Marchesini, Molinari, et al., 2013 and Compare, Calugi, Marchesini, Shonin, et al., 2013), clinical interview, and partly on patient preferences. Participants were asked to answer standardized interview questions about their last binge-eating episode, emotions preceding the binge eating episode, and whether these emotions typically preceded a binge-eating episode. EFGT was assigned whenever the consensus among the team of clinicians was that binge-eating episodes were primarily emotional in origin. CT of EFGT plus dietary counseling was assigned for those whose binge eating episodes were deemed not primarily emotional in origin or when clinicians decided that the patient with emotionally derived binge eating episodes would also benefit from comprehensive nutritional counseling. All 118 patients involved in this study received EFGT. Outcome measures were completed before the first group session and at treatment termination. Patients’ and therapists’ working alliance was assessed after the 2nd, 10th, and last (20th) group session.

The bias variable was created using a procedure described by Kenny et al. (2002). Bias scores were derived by aggregating the therapist’s ratings of working alliance scores for each the members present in a group session, excluding the therapist’s rating of the focal group member’s alliance. In this way each group member had a unique bias score for each session. Therefore, bias was a Level 1 (i.e., within-person) variable in the multilevel model. Furthermore, it is important to note that bias was a computed variable not a measured variable (i.e., based on mean therapist ratings of each group member excluding the focal group member). Group therapists did not rate the alliance of the group as a whole.

We used change scores to operationalize group member changes in binge eating symptoms, psychological distress, and obesity-related wellbeing. The change score was calculated so that higher scores represent decreases in symptoms and distress, and increases in obesity-related wellbeing. An alternative to change scores is to analyze residual scores. Fitzmaurice and Ware (2004) contend that the choice between change scores and residual scores depends on the research question. Residual scores examine an individual group member’s post treatment scores assuming that the group members start with the same scores on the variables. Change scores on the other hand, examine how much change on average happens for the group members. In this study we are interested in how much change occurred from pre- to posttreatment. We calculated both change scores and residual scores for binge eating symptoms, psychological distress and obesity-related wellbeing for the group members and found that the respective change and residual scores were highly correlated for our group members (rs > .95).

Measures

Structured Clinical Interview for DSM–IV Axis I Disorders (SCID-I/P; First, Spitzer, Gibbon, & William, 1996). The DSM–IV diagnosis of BED was based on the SCID-I/P (Compare, Calugi, Marchesini, Shonin, et al., 2013; Compare & Tassca, 2014; First, Spitzer, Gibbon, & Williams, 1996). Interviews were conducted before the therapy by intensively trained independent clinicians. These interviews were audio recorded; interviews and diagnoses were discussed in expert consensus teams that included four senior clinicians; final diagnoses were determined by consensus.
usual agreement of at least 75% of the team members. SCID-I/P interrater reliability in this study was excellent ($\kappa = 0.94$).

**Body Mass Index.** BMI was calculated by weight [kg] divided by height squared ($\text{m}^2$). Weight and height were obtained using a medical balance beam scale.

**Eating disorder examination.** Binge eating was measured during the previous 28 days based on the Eating Disorder Examination Interview (EDE; Fairburn & Cooper, 1993). The EDE was administered at baseline, and at end-of-treatment for this study. The EDE defines binge eating as consuming an unusually large quantity of food with a subjective sense of loss of control. The EDE had good inter rater reliability as assessed by intraclass correlation (ICC) coefficient ($ICC = 0.87$) among five raters in this study.

**Symptom Checklist-90 Global Severity Index.** The Symptom Checklist-90-R (SCL-90-R; Derogatis, 1994) is a 90-item measure to assess psychological symptoms and distress. The Global Severity Index (GSI) is an overall measure of psychological distress for all 90 items of the SCL-90-R. Items are rated on a 5-point scale of symptoms distress from $0 = \text{not at all}$ to $4 = \text{extremely}$. The measure has good established reliability and validity (Derogatis, 1994). For this study, the GSI coefficient $\alpha$ was 0.84.

**Obesity-Related Well-Being (ORWELL-97; Mannucci et al., 1999).** The ORWELL is an 18-item measure of obesity-related quality of life. The ORWELL has items assessing obesity-related somatic symptoms, physical functioning, and the impact of obesity on a patient’s emotional status and on her or his social relationships. Ratings are made on a 4-point scale, and higher scores reflect a clinically significant burden of obesity on quality of life. In the present study, the Cronbach’s $\alpha$ for the overall score was 0.89.

**Helping Alliance Questionnaire-II (HAq-II; Luborsky et al., 1996).** The 19-item HAq-II was used to assess working alliance. Each item is rated on a 6-point Likert scale ($1 = \text{strongly disagree}$, $6 = \text{strongly agree}$). The total score represents the quality of the working alliance, the lowest score (19) indicates absence of an alliance, and the highest score (114) indicates a very good alliance. Both the therapist and patient version of the HAq-II has demonstrated good psychometric properties (Luborsky et al., 1996). The range of Cronbach’s $\alpha$ coefficient for patients’ alliance ratings was .91 (Session 2), .89 (Session 10), and .92 (Session 20). The HAq-II forms were collected by research assistants and were not seen by the therapists.

**Treatments**

EFGT is a therapy aimed at helping clients to change the way in which they experience and use their emotions. This is achieved through an empathic and caring relationship with a therapist and group that allows one to modify personal emotion “schemes” and the self-organization responsible for problematic psychological functioning (Elliott & Greenberg, 2007). Specific goals of EFGT include: (a) identifying interpersonal needs and patterns of interpersonal behavior, (b) generating effective interpersonal behaviors to better satisfy needs, and (c) identifying and processing avoided emotions associated with all therapeutic content (Elliott & Greenberg, 2007). In EFGT, the therapist helps clients to reflect on and reevaluate emotion schemes linked with binge episodes, become aware of their emotional reactions, and expose themselves to more adaptive emotional responses. Moreover, therapists helped clients to access subdominant adaptive emotional responses. Research examining EFT in a group setting is scarce (Pascual-Leone, Berman, Arnold, & Stasiak, 2011), but recent reports showed it is a promising intervention for binge eating (Compare, Calugi, Marchesini, Shonin, et al., 2013; Compare & Tasca, 2014). In this study, EFGT was administered via 20 group therapy sessions (10–18 participants per group with weekly sessions of 60–90 min), in eight groups over a 5-month treatment period. Approximately half of the patients received EFGT combined with dietary counseling (i.e., CT, see Procedure). In the current study, the dietary counseling, which aimed to lower energy dense food intake, was administered via 12 1-hr weekly individual sessions over the first 3-months of treatment by clinicians different from the group therapists. Clinicians provided information on sample meals, menus, and recipes; reviewed the weekly topic with patients; and helped to remove possible obstacles to achieve the goal of lower energy density. Participants were asked to complete daily food diaries that were checked weekly by clinicians.

Eight group therapists were involved in the study, and each therapist led one group. All therapists participated in 2-year (part-time) postgraduate training program with an EFGT focus and had at least 1 year of training before beginning to see clients. The therapists received 6 months supervision to be adherent to the EFGT for BED manual (Compare & Molinari, 2012). Therapists provided groups in both treatment conditions.

**Results**

The means and SDs for the variables examined in the T&B analysis are displayed in Table 1. Initially we ran two 3-level hierarchical linear model (HLM) base models to partition the variance into group therapist-rated and group member-rated working alliance. The therapist-level (Level 3) ICC for group therapist-rated working alliance was .01 ($p = .275$), and for group member-rated working alliances ICC = .01 ($p = .391$). Nevertheless, we specified three-level models because the significance tests for the ICCs were underpowered.

Initially we ran a three-level T&B model with no moderator variables. The specific HLM model examined is displayed in supplementary online material. The results of the T&B analysis are displayed in Table 2. These results will be reported in relation to our hypotheses.

Because of the contradictory findings in the literature we did not make a hypothesis concerning the truth effect. Table 2 shows that the truth effect was not significant, $\gamma_{300} = -0.09$, $p = .405$. Therefore, there was no significant relationship between group therapists’ and group members’ working alliance perceptions. As hypothesized, however, there was a significant bias effect, $\gamma_{500} = 1.18$, $p = .002$. The $d$ statistic shows that for every one unit change in the bias variable (therapists’ perceptions of other group members’ alliances) there was a .77 $SD$ change in the judgment variable. Therefore, when the group therapists saw their alliances with the other group members’ as strong they also saw their alliance with a specific group member as strong. Also as hypothesized, there was a significant bias by time interaction, $\gamma_{500} = -0.30$, $p = .013$. The $d$ statistic shows that for every one unit change in this
interaction variable there was a $-0.20$ SD change in the judgment variable.

The form of the bias strength by time interaction is displayed in Figure 1. A simple slopes analysis revealed that the bias slope in early sessions was positive and significant ($b = 3.42, p < .001$), whereas the bias slope in later sessions was not significant ($b = -1.01, p > .05$). This means that group therapists' ratings of an individual group member were significantly related to their ratings of the other group members in early sessions but unrelated to their ratings of other group members in later sessions.

As hypothesized, the directional bias coefficient was negative and significant, $\gamma_{100} = -1.25, p = .027$. The group therapists underrated the alliances as suggested by the better safe than sorry hypothesis. Finally, contrary to our hypothesis the interaction between truth strength and time was not significant, $\gamma_{400} = 0.07, p = .345$.

There was significant between therapist variance in directional bias (variance component = 1.14, $p < .001$), bias strength (variance component = 0.44, $p < .001$), and in the time by bias strength interaction (variance component = 0.05, $p < .001$) but not in truth strength (variance component = 0.01, $p > .500$) or in the truth strength by time interaction (variance component = 0.00, $p > .500$). Therefore, some group therapists had more discrepancy and some therapists had less discrepancy in their ratings of their group members' alliances. Also for some group therapists there was greater bias strength than for other therapists. Finally, some group therapists had greater reductions in bias strength over time than did other group therapists. The moderator hypotheses were examined with a three-level conditional HLM. This specific model is displayed in supplementary online material.

Contrary to our hypothesis there were no significant cross-level interactions between any of the change variables and the directional bias intercept (see Table 3). Also contrary to our hypothesis there were no significant cross-level interactions between any of the change variables and truth strength (see Table 3). As hypothesized, there was a significant cross-level interaction between changes in days binged and bias strength, $\gamma_{320} = -0.09, p = .029$. The $d$ statistic shows that for every one unit change in this interaction variable there was a $-0.25$ SD change in the judgment variable (see Table 3).

The form of this interaction is displayed in Figure 2. A simple slopes analysis revealed that the slope for group members who made greater changes in the number of days binged (1 SD above the mean) was not significantly different from zero, $b = 0.23, p > .05$. The slope for group members who made fewer changes in the number of days binged (1 SD below the mean) was positive and significant ($b = 0.63, p < .05$). For members who had greater improvement in binge eating: how their group therapist perceived that group member’s working alliance was unrelated to the working alliances of the other group members. Conversely, for members who made less improvement in binge eating: how their group therapist perceived that group member’s working alliances was positively and significantly related to the working alliances of the other group members.

**Discussion**

The current study was the first group therapy study to use the T&B model to examine therapists’ ability to accurately perceive their clients’ alliances. According to the T&B model, we treated

**Table 1**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group therapist working alliance early</td>
<td>4.55</td>
<td>1.31</td>
<td>2.13</td>
<td>6.00</td>
</tr>
<tr>
<td>Group therapist working alliance-middle</td>
<td>2.20</td>
<td>0.90</td>
<td>1.05</td>
<td>4.87</td>
</tr>
<tr>
<td>Group therapist working alliance-late</td>
<td>2.23</td>
<td>0.97</td>
<td>1.05</td>
<td>5.34</td>
</tr>
<tr>
<td>Group member working alliance early</td>
<td>2.76</td>
<td>1.00</td>
<td>1.15</td>
<td>5.96</td>
</tr>
<tr>
<td>Group member working alliance-middle</td>
<td>2.12</td>
<td>0.78</td>
<td>1.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Group member working alliance-late</td>
<td>3.56</td>
<td>1.38</td>
<td>1.05</td>
<td>5.34</td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in Global Severity Index</td>
<td>1.23</td>
<td>1.23</td>
<td>-1.50</td>
<td>5.00</td>
</tr>
</tbody>
</table>

**Note.** Group member $N = 118$; session $N = 354$. 

**Table 2**

<table>
<thead>
<tr>
<th>The Unconditional Truth-and-Bias Model for Group Leader Alliance Ratings</th>
<th>Coefficient (d)</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directional bias, $\gamma_{100}$</td>
<td>$-1.26 (-0.93)$</td>
<td>.45</td>
<td>-2.78</td>
<td>.027</td>
</tr>
<tr>
<td>Time (early, middle, or late), $\gamma_{100}$</td>
<td>$-1.10 (-0.07)$</td>
<td>.02</td>
<td>-5.60</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Truth (group member working alliance), $\gamma_{200}$</td>
<td>$-0.09 (-0.06)$</td>
<td>.10</td>
<td>-8.9</td>
<td>.405</td>
</tr>
<tr>
<td>Bias (aggregated other member working alliance), $\gamma_{300}$</td>
<td>$1.18 (0.77)$</td>
<td>.26</td>
<td>4.60</td>
<td>.002</td>
</tr>
<tr>
<td>Time by truth interaction, $\gamma_{400}$</td>
<td>$0.01 (0.01)$</td>
<td>.01</td>
<td>1.01</td>
<td>.345</td>
</tr>
<tr>
<td>Time by bias interaction, $\gamma_{500}$</td>
<td>$-0.30 (-0.20)$</td>
<td>.09</td>
<td>-3.29</td>
<td>.013</td>
</tr>
</tbody>
</table>

**Note.** Group member $N = 118$; session $N = 354$. 

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
group members’ ratings of the working alliance with their group therapist as the truth, the group therapists’ ratings of this working alliance with each of their group members as “judgments” about the member’s working alliance, and the aggregated alliance ratings that the therapist made for the other group members as bias.

We did not make a hypothesis about truth strength because previous studies found mixed results concerning group therapists’ ability to accurately perceive group process. Our results for truth strength are similar to Jenkins et al. (1971) who did not find a significant congruence between group therapists’ and group members’ views of the therapeutic relationship. Our results support Chapman et al.’s (2012) conclusion that group therapists are generally not accurate in their perceptions of how group members perceive group therapeutic relationships.

As hypothesized, the group therapist in our study appeared to adopt a better safe than sorry strategy (Atzil-Slonim et al., 2015) when it came to judging the strength of their group members’ alliances. As in individual therapy, EFGT therapists erred on the side of seeing the alliance as weaker than it really was. As noted above, the group setting necessitates different alliances, which occur at different levels (Kivlighan & Kivlighan, in press). There are many more alliances for the therapist to track in the group setting which places an increased cognitive demand on the therapist and can result in resource depletion. Based on the current finding we could speculate that it may be much more difficult for group therapists to accurately perceive their alliances with individual group members than it is for individual therapists to accurately perceive their alliance with a specific client. Therefore, EFGT therapists in this study may have adopted the better safe than sorry strategy, because individual alliances in groups are difficult to perceive accurately.

We hypothesized that group therapists may reduce the cognitive demands of keeping track of so many alliances by using a strategy

Table 3

<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>Coefficient (standardized)</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directional bias, γ_{010}</td>
<td>-4.95 (−3.28)</td>
<td>1.18</td>
<td>-4.22</td>
<td>.004</td>
</tr>
<tr>
<td>Change in BMI, γ_{010}</td>
<td>−0.02 (−.01)</td>
<td>.02</td>
<td>-67</td>
<td>.525</td>
</tr>
<tr>
<td>Change in days binged, γ_{020}</td>
<td>−0.02 (−.02)</td>
<td>.03</td>
<td>-.79</td>
<td>.454</td>
</tr>
<tr>
<td>Change in ORWELL-97, γ_{030}</td>
<td>.01 (.01)</td>
<td>.01</td>
<td>1.01</td>
<td>.347</td>
</tr>
<tr>
<td>Change in GSI, γ_{040}</td>
<td>.02 (−.01)</td>
<td>.04</td>
<td>.45</td>
<td>.666</td>
</tr>
<tr>
<td>Time (early, middle, or late), γ_{050}</td>
<td>−0.63 (−1.8)</td>
<td>.19</td>
<td>-3.29</td>
<td>.013</td>
</tr>
<tr>
<td>Truth (group member working alliance), γ_{200}</td>
<td>.02 (−.01)</td>
<td>.10</td>
<td>.18</td>
<td>.866</td>
</tr>
<tr>
<td>Change in BMI, γ_{210}</td>
<td>.01 (.01)</td>
<td>.02</td>
<td>.70</td>
<td>.504</td>
</tr>
<tr>
<td>Change in days binged, γ_{220}</td>
<td>−0.01 (−.01)</td>
<td>.01</td>
<td>-.988</td>
<td>.356</td>
</tr>
<tr>
<td>Change in ORWELL-97, γ_{230}</td>
<td>−.01 (−.01)</td>
<td>.01</td>
<td>-.102</td>
<td>.921</td>
</tr>
<tr>
<td>Change in GSI, γ_{240}</td>
<td>−.01 (−.01)</td>
<td>.03</td>
<td>-1.17</td>
<td>.873</td>
</tr>
<tr>
<td>Bias (aggregated other member working alliance), γ_{300}</td>
<td>.49 (.28)</td>
<td>.87</td>
<td>.56</td>
<td>.594</td>
</tr>
<tr>
<td>Change in BMI, γ_{310}</td>
<td>−.03 (−.02)</td>
<td>.03</td>
<td>-1.16</td>
<td>.284</td>
</tr>
<tr>
<td>Change in days binged, γ_{320}</td>
<td>−.09 (−.27)</td>
<td>.03</td>
<td>-2.67</td>
<td>.029</td>
</tr>
<tr>
<td>Change in ORWELL-97, γ_{330}</td>
<td>.01 (.01)</td>
<td>.02</td>
<td>.99</td>
<td>.572</td>
</tr>
<tr>
<td>Change in GSI, γ_{340}</td>
<td>.02 (.01)</td>
<td>.05</td>
<td>.422</td>
<td>.685</td>
</tr>
<tr>
<td>Time by truth interaction, γ_{40}</td>
<td>.00 (.01)</td>
<td>.01</td>
<td>.143</td>
<td>.890</td>
</tr>
<tr>
<td>Time by bias interaction, γ_{50}</td>
<td>−.96 (−.62)</td>
<td>.08</td>
<td>-11.99</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note. Group member N = 118; session N = 354. BMI = Body Mass Index; ORWELL = obesity related quality of life; GSI = Global Severity Index.
of perceiving their alliance with a specific group members as being similar to their alliances with the other members of the group, rather than perceiving the group members as having unique alliances. The results regarding bias strength supported this hypothesis. There was a strong relationship between the alliance ratings of the other group members and the alliance ratings of the individual group member. This suggests that the group leaders in EFGT did not distinguish among the group members in terms of distinctive strengths of their alliances. In other words, EFGT therapists appeared to have had a general perception of the alliances of the other members in the group, and this general perception was applied to an individual alliance.

On the other hand, group therapists may accurately perceive that a group member’s alliance is related to the other group members’ alliances because there was a group nesting effect for member-rated alliances. However, if the relationship between group therapists’ rating of a group member and their ratings of the other group members’ was an accurate perception of the within-group dependency of alliance scores then this accuracy should be related to a better outcome for the member, which was not the case. It is likely that part of the significant bias strength found in this study represents therapists’ accurately accounting for within-group dependency in member-rated alliances, and therapists inaccurately rating the strength of the within-group dependency. Our results suggest that when group therapists misjudge the level of within-group dependency for a particular group member, that group member has less change in binge eating symptoms.

Based on Hamilton and Sherman’s (1996) studies of how information about individuals and groups of individuals is stored, we suggested that therapists use an aggregated perception of the other group members’ alliances as a proxy for each individual’s alliance as a way to conserve their cognitive resources. Our results suggest that group leaders in EFGT do attempt to prevent resource depletion by perceiving group member alliances as an aggregate. These therapists appeared to use this perceived aggregate alliance of the other group members in part to represent the alliance of each of the group members.

Also hypothesized, bias strength decreased over time. As the EFGT groups developed and group members became more differentiated and individualized, the group leaders were less likely to confound their judgments of individual alliances with their aggregate judgments. This suggests that this aggregation bias may have a self-limiting quality, such that group leaders are increasingly able to differentiate alliances with individuals over time. This is consistent with Chapman et al. (2012) who found that therapist’s predictive accuracy increased in later sessions for counseling center groups. Therefore, it is especially important for group leaders to have strategies to overcome this aggregation bias in early group sessions.

We also hypothesized and found that there was a weaker bias effect for members who benefited the most from the groups in terms of reduced binge eating. As suggested above, the “default” position for group leaders is to perceive a group member’s alliance in terms of the group’s aggregate alliance. When a group leader overcomes this default strategy and sees a group members’ alliance in a less biased manner then that group member derives more benefit. In the context of EFGT for BED, if the group therapist accurately perceives that the group member has a weak alliance, the group therapist can attempt to strengthen the group member’s alliance. Conversely, accurately perceiving that a group member has a strong alliance can afford the group therapist the opportunity to more deeply work on interpersonal problems that lead to emotion deregulation that may underlie the binge eating.

According to Funder (1995) accurate perceptions occur when there is a good target, good judge, good trait, and good information. Alliances may not be especially good or easy traits to judge; therefore, we suggest that future group research should examine what makes group therapists “good judges,” what makes group members “good targets,” and what allows group sessions to provide “good information.” Given the cognitive demands of perceiving alliances in group therapy, variables like cognitive complexity or flexibility may be assessed by researchers for their potential impact on therapists’ ability to accurately rate the alliance.

Furthermore, individual treatments sessions in which the client was more distressed seemed to provide “bad information” for alliance accuracy (Atzil-Slonim et al., 2015). Atzil-Slonim et al. (2015) speculated that therapists might misperceive their clients’ distress as signs of alliance ruptures leading to inaccurate alliance judgments. Flowers, Booraem, and Seacat (1974) found that group therapists more accurately perceived group processes in sessions during which the group therapist received negative feedback. Per-
haps group therapists are more attuned to the therapy relationship when they feel challenged.

Research in individual treatments has not yet identified characteristics of good targets in relation to accurate alliance perceptions (Atzil-Slonim et al., 2015). One could speculate that self-disclosure may make a group member a better target. The more group members disclose the more they differentiate themselves from other members, allowing the group therapist to make more accurate judgments.

Finally we found that group therapists’ ratings of the alliance with their group members significantly decreased across the three measurement occasions. We are not aware of any studies that examined how therapist-rated alliance to group members changes across time in group therapy. However, in individual treatments therapist-rated alliances either increased linearly across sessions (Kivlighan & Shaughnessy, 1995) or followed a dual slope model with a steeper linear slope earlier in treatment and a shallower linear slope later in treatment (Chu, Skriner, & Zandberg, 2014). The decreasing pattern of group therapist-rated alliance in the current study is clearly different from the therapist-rated alliance patterns reported in individual treatments. Group therapists’ average alliance ratings were over a scale point higher than average client ratings early in EFGT and were over a scale point lower than average client ratings late in EFGT. Only during midtreatment were mean therapist and client alliance ratings congruent. Group therapists may have been adjusting their early high alliance ratings to be more congruent with their group members’ alliance ratings, but they appeared to “overshoot” their adjustment by the end of treatment.

Because group therapist alliance ratings was the criterion variable in this study we did not conduct a formal statistical test of how group member alliance ratings changed over time. Examining the means (see Table 1), however, suggests that group member-rated alliance followed a high-low-high pattern, which is consistent with previous research focusing on group climate (McClendon & Burlingame, 2010). To our knowledge, there are no studies of group member alliance that have documented this type of high-low-high alliance development pattern. Tasca et al. (2010) found a linear increasing pattern of group member alliance ratings; whereas Lorentzen, Bakali, Hersoug, Hagtvet, Ruud, and Høglend (2012) found no significant increase or decrease in group member-rated alliance across time. Other group relationship variables like engagement (i.e., cohesion) do show a high-low-high pattern of development (e.g., Bakali, Wilberg, Klungsøyr, & Lorentzen, 2013). We need more research on how group leader-rated and group member-rated alliance develops over time.

**Strengths and Limitations**

The T&B model provides advantages over simply examining group member—group therapist correlations because: (a) it allows for a simultaneous examination of directional differences and of correlations in group member and group leader perceptions; (b) it allows researchers to incorporate bias variables into an accuracy model; and (c) it easily accommodates moderators of directional bias, truth strength, and bias strength. Furthermore, this was only the second study, following Atzil-Slonim et al. (2015) to examine alliance accuracy longitudinally. A longitudinal examination of alliance perceptions is especially important because studies repeatedly show that most of the variance in alliance perceptions is at the session (i.e., within-person) level (Gullo et al., 2014). On the other hand, there is a literature that suggested that there is a considerable amount of variance in the therapeutic relationship at the group level. For example, Thayer and Burlingame (2014) found Group ICCs accounting for 28% of the variance on the therapeutic relationship. Further research should continue to address the within-group dependent issue in the analysis of group data (Baldwin, Stice, & Rohde, 2008).

There are also several limitations that must be considered when interpreting the study’s results. First, we examined a relatively small number of groups and therapists in this study. Future research that attempts to identify aspects of group therapists that would make them a good judge will need to include more group therapists. Second, the data came from a study with strict inclusion criteria, thorough group therapist training, and specific manualized treatment protocols. Therefore, it is not clear how well our results will generalize to usual group therapy practiced in the community. Third, we examined a specific type of therapy (EFGT) with a specific patient population (obese individuals with BED) and so the findings need to be replicated with other therapy types and patient populations. Fourth, we only examined group member to group leader alliances in this study. However, alliances in group treatments exist at several different levels (e.g., member to member alliance; Tasca, Francis, & Balfour, 2014). Therefore, we do not know how accurately group therapists can judge alliances at these other levels, and the congruence among group members in their judgment of the alliance. Finally, alliance ratings were only made at three points in time. Therefore, we were limited to assessing linear changes in therapist rated group member alliances and in truth and bias strength. As noted above, no research has examined how group leaders’ or group members’ alliances change over time. Future research should examine more complex growth models describing how group therapists’ and group members’ alliances change over time.

**Conclusions**

The current data inform our understanding of group therapists’ ability to accurately perceive their clients’ alliance and its link with treatment outcome. In general group therapists appear to use their perception of the group’s aggregate alliance to judge each group member’s individual alliance. However, when the group therapist overcomes this general tendency with a particular group member that member has a more successful outcome. Group therapists may need to make more fine-grained distinctions among the alliances they have with their individual group members. Furthermore, because individual group member alliances may be hard for group therapists to judge accurately, group therapists should consider collecting and examining group member alliance data on an ongoing basis (Reese, Slone, & Misericocchi, 2013). For example, the Group Questionnaire (GQ) is a measure that is repeatedly administered to group members over the course of treatment to track patient change and dimensions of the therapeutic relationship (Krogel et al., 2013). Such measures can be implemented in evidence-based feedback systems (Lambert & Lo Coco, 2013), which reduce premature termination and improve outcomes in individual therapy (Lambert & Shimokawa, 2011) and group therapy (Davies, Burlingame, Johnson, Gleave, & Barlow, 2008). For
example, in the current study some patients reported a worsening in their BMI and the SCL-90 GSI scores (see Table 1). A feedback to therapist addressing these poor patients’ alliance scores over the course of treatment could have helped therapists to better address the weakness of therapeutic relationship with these nonresponders.

The findings of the current study also suggest that this type of feedback system could help group therapists to counter their tendency to see group member alliances as an aggregate, and therefore therapists can more effectively help their individual group members. Therapists who accurately judge their clients to have a weaker alliance can work at improving the alliance, whereas therapists who accurately judge their clients as having a stronger alliance can challenge that client to do deeper therapeutic work.

References


