Psychological Bulletin
2010, Vol. 136, No. 4, 627–658

0033-2909/10/$12.00 DOI: 10.1037/a0019792

Through the Eyes of Love: Reality and Illusion in Intimate Relationships

Garth J. O. Fletcher and Patrick S. G. Kerr
University of Canterbury

This article reviews the research literature and theory concerned with accuracy of judgments in romantic relationships. We initially propose a model of cognition in (romantic) relationships that distinguishes between 2 forms of accuracy: mean-level bias and tracking accuracy. We then report the results of meta-analyses of research on heterosexual, romantic relationships, which used external benchmarks and reported levels of tracking accuracy (98 studies) and/or mean-level bias (48 studies). The results revealed robust overall effect sizes for both tracking accuracy ($r = .47$) and positive mean-level bias ($r = .09$). As expected, the effects were substantial and positive for tracking accuracy across 6 judgmental categories, whereas signed mean-level bias was negative for the interaction attributions (e.g., love, communication). The results showed, as expected, that these 2 forms of accuracy were independent—the 2 kinds of effect size derived from the same set of 38 studies were uncorrelated. As expected, gender, relationship length, and relationship evaluations moderated mean-level bias across studies but (unexpectedly) not for tracking accuracy. In the Discussion we evaluate the prior model in light of the findings, other research, moderating variables (such as self-esteem), the role of projection, the early stages of mate selection, metacognition, and the rationality and nature of motivated cognition. We conclude that our model, findings, and analyses help to resolve the apparent paradox that love is both riven with illusions and rooted in reality, and support both evolutionary and social psychological approaches to understanding cognition in romantic relationships.

Keywords: bias, accuracy, romantic relationships, meta-analysis

> Love looks not with the eyes but with the mind; And therefore is wing'd Cupid painted blind. —William Shakespeare

Romantic love is often characterized as shot through with illusions and driven by strong emotions and wishful thinking—as "painted blind," to use Shakespeare's phrase. This thesis is theoretically plausible. From an evolutionary standpoint, romantic love is typically viewed as a commitment device designed to lead men and women to substantially invest for long periods in one another, and accordingly support any resultant offspring (Gonzaga & Haselton, 2008). The leap of faith required for long-term romantic commitments is thus likely to be powered by strong biologically based attachment emotions, which, in turn, predispose individuals to put a charitable spin on judgments of their partners and relationships.

The theoretical credibility of this thesis is supported by a wealth of empirical evidence. For example, people routinely rate the chances of their own marriages failing as considerably less than their perceptions of the population base rates (Fowers, Lyons, Montel, & Shaked, 2001), and keep their relationship doubts at bay by restructuring judgments or rewriting their relationship stories (see Murray, 2001). And as love prospers and grows more intense, individuals increasingly exaggerate the similarity with their partners (Murray, Holmes, Bellavia, Griffin, & Dolderman, 2002), the extent to which their relationships have improved over time (Karney & Frye, 2002), and the extent to which their real-life partners resemble archetypal ideals (Murray, Holmes, & Griffin, 1996a, 1996b).

There are, however, strong arguments against the love-is-blind hypothesis. The simple fact that many long-term romantic relationships dissolve suggests that the motivating power of love to promote positive bias has its limitations. In addition, the use of evolutionary psychology to support this thesis is a double-edged sword. Evolutionary psychology rests on the Darwinian assumption that mate selection criteria in species (including humans), and their associated attributes, evolved according to natural and sexual selection. The force of sexual selection to produce the cumbersome, dazzling tail of the peacock relies on the ability of the peahen to perceive the size and quality of the peacock's tail accurately. What typically goes unnoticed is that this Darwinian assumption also entails that human judgments of important attributes in the mating game, such as attractiveness, status, and kindness or trustworthiness, are assessed with reasonable accuracy. If, instead, such judgments of potential or actual partners are (or were in our ancestral environment) hopelessly awry, then it is hard to see how their use would have played a role in the development or evolution of such attributes. Thus, the love-is-blind thesis, taken to extremes, undercuts a key assumption in evolutionary psychology.

Moreover, a broad array of empirical evidence suggests that lay judgments of partners and relationships are firmly tied to reality. For example, relationship evaluations quite strongly predict both interactive behavior (e.g., Fletcher & Thomas, 2000) and relation-

Garth J. O. Fletcher and Patrick S. G. Kerr, Psychology Department, University of Canterbury, Christchurch, New Zealand.

Thanks are extended to David Kenny, Kim Nathan, and Jeffry Simpson for their helpful comments on a prior draft.

Correspondence concerning this article should be addressed to Garth J. O. Fletcher, Psychology Department, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. E-mail: garth.fletcher@canterbury.ac.nz

ship dissolution (see Karney & Bradbury, 1995), and are shared across partners (e.g., Campbell, Simpson, Kashy, & Fletcher, 2001). Further, studies using a range of external criteria or benchmarks (including self-reports of the partner, observer ratings of interactive behavior, and the predicted future or recalled past states of the relationship) reveal quite good levels of accuracy in relationship and partner judgments (for a recent review, see Fletcher & Boyes, 2008).

Thus, we are left with an apparent paradox: Love is both blind and firmly rooted in the real world. Few (if any) psychologists have argued that love is completely free of reality constraints or that judgments in intimate relationships are not biased in various ways. This observation reasonably suggests that both sides of the argument have some validity, but still leaves the conundrum unsolved. In this article we offer some solutions and, in the process, draw some conclusions about how lay social cognition functions in intimate relationship contexts.

Few of the key ideas put forward here are novel, and previous reviews have dealt with the same domain (e.g., Fletcher & Boyes, 2008; Gagné & Lydon, 2004). However, prior reviews have been qualitative and partial in nature, with no published meta-analyses of the subsequent findings. This article first presents a theoretical and methodological analysis, then a related systematic meta-analysis of findings in the intimate relationship domain. This undertaking should answer some difficult questions, help establish the facts, assist in building theory, and generate further predictions and questions.

As noted by Fletcher and Boyes (2008), research and theory dealing with bias and accuracy in intimate relationships have largely emerged in the context of more general scientific work, rather than appearing out of the blue. Indeed, the conventional wisdom that everyday social cognition is generally rational and designed to detect the truth about ourselves and others has taken a pounding over the last 4 decades. Doubts about the validity of lay social cognition reached their zenith in the 1980s, with bold claims being advanced that it was inherently flawed or even irrational (Nisbett & Ross, 1980). Extensive research in social cognition apparently showed that people are subject to many biases, including the fundamental attribution error, the false consensus effect, the confirmation bias, and self-serving biases (for reviews, see Fletcher, 1995; Krueger & Funder, 2004). Moreover, an influential review by Taylor and Brown (1988) argued that an important subset of these biases—including the tendency to adopt Pollyannaish self-perceptions and exaggerate self-control—is associated with superior mental health, personal happiness, and healthy relationships.

Such jaundiced views of common-sense social cognition, perhaps inevitably, provoked a rash of research and argument seeking to show that lay social cognition is not as bad as has been claimed, that lay social judgments are surprisingly accurate much of the time, that ignorance is not bliss, and that seeking truth in social life does not doom people to sad lives and impoverished relationships (for examples of reviews that strike these kinds of themes, see Colvin & Block, 1994; Fiedler & Wänke, 2009; Fletcher, 1995; Funder, 1995; Klayman & Ha, 1987; Krueger & Funder, 2004). Thus, it can be seen that arguments in the relationship arena mirror to some extent the debate in the general (social) cognitive and personality domain.

The first research investigating accuracy in intimate relationships that we could find was published in 1954 by Dymond (in this case studying personality judgments in married couples), and Kahn (1970) published the first study of mind reading in marriages 40 years ago. Both studies used techniques that became the forerunners of later work. Dymond had 15 married couples rate whether personality traits were either true or false for both self and partner. Kahn pioneered a technique in which married partners read statements out to one another (e.g., "Didn't we have chicken for dinner a few nights ago?") with instructions to use nonverbal cues to indicate either negative (irritation), neutral (curiosity), or positive (elation) emotional reactions. The perceivers' task was to assess what their spouse was intending to communicate. In both research cases, the authors concluded that those in happier marriages were more accurate, or, as Dymond concluded, "married love is not blind, and ignorance is not connubial bliss" (p. 171).[1] However, relationship researchers on both sides of the argument entered the fray with gusto from about 20 years ago.[2]

It perhaps goes without saying that an analysis of research in this area should help resolve some key questions in understanding relationship cognition, including the conundrum outlined previously. However, work in this domain also offers a useful case study for helping to build a general understanding of bias and accuracy in social judgments, given that the motivating forces stacked up on both sides of the ledger—for positive bias and accuracy, respectively—seem especially strong in intimate romantic relationships. Understanding when, why, and how laypeople resolve the two kinds of associated motivational sets in intimate relationship contexts has implications for the wider study of motivated cognition and rationality in lay social cognition.

First, we present an overall analysis that distinguishes between two kinds of accuracy in social judgment: mean-level bias and tracking accuracy. Second, we briefly discuss two relevant, but quite different, kinds of bias that are important in the study of accuracy in social judgments (belief perseverance and projection). Third, we present a general model of social cognition in intimate relationships as a springboard for our predictions and questions, both for the meta-analysis and as an explanatory and integrative tool for the wider discussion of research in the Discussion section (see Figure 1). Fourth, we derive the predictions and questions posed in the subsequent meta-analyses of the relevant research literature. Finally, after reporting the results of the meta-analysis, we discuss the implications for our understanding of intimate relationships, and draw on a wider literature to discuss some issues not dealt with in the meta-analysis. These latter issues concern the role of different moderators (such as individual differences), the study of projection in relationships, accuracy in the early stages of mate selection, the extent to which people are aware of illusions or

[1] The study by Kahn (1970) was subsequently used in the meta-analysis, but the Dymond (1954) study was not usable because of methodological deficiencies and lack of information to calculate effect sizes.

[2] Exemplifying this point, of the 112 research articles identified for our meta-analyses, 88% were published in the last 20 years and only three were published prior to 1980. Interest in this topic, perhaps carried along by the surge of interest in intimate relationships over the last decade or so, also appears to be accelerating, with 33% of our sample of research articles being published in the 5-year period from 2004 to 2008.

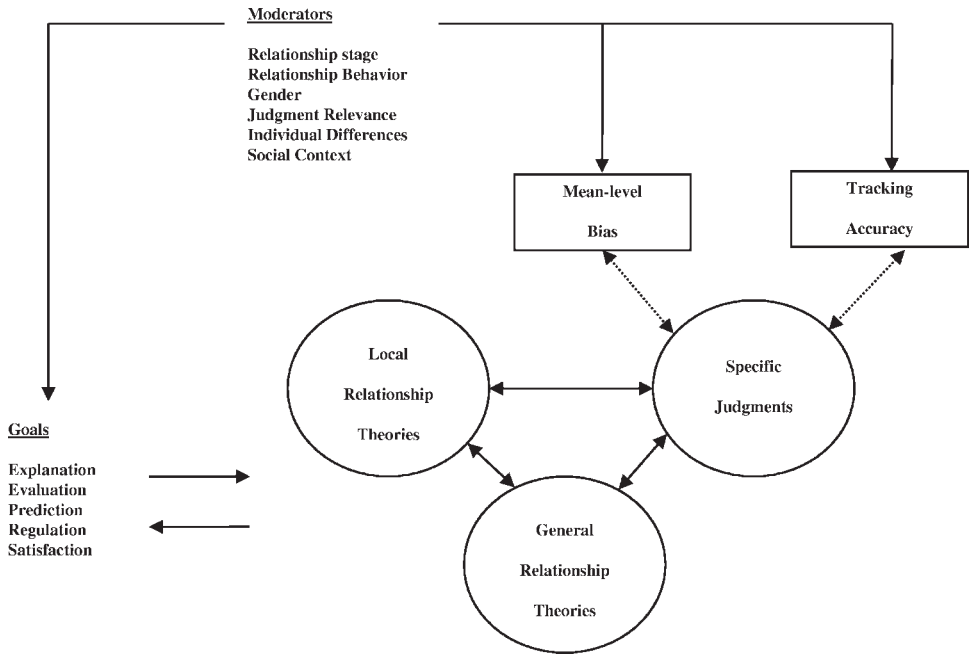

Figure 1. General model of relationship cognition showing links between mean-level bias and tracking accuracy of specific judgments with relationship goals, general and local relationship theories, and moderators (solid lines represent causal links, and dotted lines indicate qualities of specific judgments).

verities in their own relationships, and the implications for the wider study of motivated cognition and the rationality of lay social cognition.

Two Kinds of Accuracy in Intimate Relationships: Mean-Level Bias and Tracking Accuracy

This section lays the groundwork for the subsequent literature review. We initially define the accuracy of judgments simply in terms of their correspondence with reality. But as psychologists have long known, the empirical investigation of accuracy quickly turns such an apparently straightforward definition into a sea of complexity. A fundamental beginning point is to distinguish between two kinds of accuracy: mean-level bias and tracking accuracy. This distinction is not new, having been pointed out by several authors over the years both in the relationship domain (Fletcher, 2002; Gagné & Lydon, 2004) and in other areas of research (Cronbach, 1955; Epley & Dunning, 2006; Fletcher, Reeder, & Bull, 1990; Funder & Colvin, 1997; Kenny & Albright, 1987).[3] However, as noted by Epley and Dunning (2006), its importance has largely gone unrecognized, and the distinction is often blurred or ignored in empirical articles.

Mean-level bias can be defined as differences in mean levels across a sample of a specific judgment comparing the judgment with some benchmark. *Tracking accuracy* can be formally defined as the association (typically a correlation) between a judgment or set of judgments and a relevant benchmark or set of benchmarks.

Consider the following example (adapted from Fletcher, 2002). As shown in Table 1, imagine that for a group of four couples, each man rates his female partner on the trait of warmth, and that we have gold standard criteria showing that, in reality, Mary is 5 on warmth, Joanne 4, Iris 4, and Anna 3 (all on 7-point scales). If we take the first column of possible judgments (Set a), it can be seen that all four male partners score a direct hit. Thus, the sample overall would produce no mean-level bias and attain perfect tracking accuracy. The next pattern of male ratings (Set b) shows an elevated pattern of positive bias, but because each man rates his female partner at two units above her actual level of kindness, the sample shows perfect tracking accuracy ($r = 1.0$). Following the same logic, the remaining two sets (c and d) show patterns of low tracking accuracy and high positive bias, and low tracking accuracy and no bias, respectively.

In this example, tracking accuracy is assessed on one item (warmth) across the sample. Another method sometimes used involves assessing correlations at the within-couple level, which of necessity uses more than one trait. With this method, the overall level of tracking accuracy attained at the sample level is the mean of the within-couple correlations. These methods differ in various technical ways (see Kenny & Winquist, 2001), which we mention later on. However, regardless of which method is used, such correlational methods produce an index of relative accuracy that does not take into account mean-level differences. This can clearly be seen, for example, in the first two columns of Table 1 showing

[3] Several terms have been used to signal this distinction, including discrimination, bias, and elevation accuracy, for what we term mean-level bias, and resolution, accuracy, and calibration, for what we term tracking accuracy. We prefer the terms we use here because we think their meanings are clearer and more technically accurate than the prior terms employed. The other major way of assessing accuracy is in terms of consensus judgments about a target, which is rarely done in relationship research.

Table 1
An Example Illustrating the Potential Independence of Mean-Level Bias and Tracking Accuracy of Partner Judgments Across One Trait in a Sample of Four Couples

Benchmark rating	(a) High accuracy, no bias	(b) High accuracy, high bias	(c) Low accuracy, high bias	(d) Low accuracy, no bias
Female targets				
Mary (5)	5	7	6	2
Joanne (4)	4	6	7	2
Iris (4)	4	6	5	6
Anna (3)	3	5	6	6
Mean-level bias[a]	4	6	6	4
Tracking accuracy[b]	1.0	1.0	.00	−.71

Note. Benchmark ratings are shown in parentheses after each name. The above example assumes a scale of 1 to 7, where 7 = *warm* and 1 = *cold.*
[a] Mean rating across group. [b] Correlations between perceptions and benchmarks across group.

that the male partners' judgments in both cases produce correlations of 1.0, yet in the second column (high tracking accuracy and high mean-level bias) the men are systematically erring in their judgments.

It should also be noted that the effect sizes for mean-level bias and tracking accuracy are not equivalent in terms of measuring overall accuracy. For mean-level bias, an effect size of zero represents no bias (thus maximum accuracy), and mean effect sizes of −.40 and .40 represent equal amounts of bias (negative and positive, respectively). In contrast, effect sizes for tracking accuracy theoretically range from −1.0 to 1.0. In practice, in our review we found no cases in which tracking accuracy was negative for the sample; thus, a correlation of zero can be considered as representing the bottom end of the scale for this construct. However, we did find (as expected) some cases in which the average mean-level bias across a sample was negative. Thus, where it was appropriate and possible to do so, we calculated two kinds of mean-level bias: *positive* mean-level bias, as in the prior examples, and *absolute* mean-level bias, which counts all effect sizes as positive (and simply represents the amount of nonsigned bias). Almost all the theoretical treatments, and associated research findings, are interpreted in terms of positive mean-level bias. Thus, we concentrate on this kind of calculation. Nevertheless, both kinds of measure can be informative, as shall be seen.

In terms of evaluating the overarching levels of accuracy attained by the sample, we suggest that both mean-level bias (positive or absolute) and tracking accuracy should be interpreted as components of overall levels of accuracy. Indeed, as the example just given shows, the logic seems unassailable on this point. Maximum overall accuracy in Table 1 is attained in the unbiased and accurate tracking condition, with the other three exemplars illustrating various degrees of inaccuracy. However, the fundamental point remains fast that these two criteria of accuracy are potentially independent across samples. Whether they are actually independent is an empirical question and one that our meta-analysis deals with. We continue to use the terms *mean-level bias* and *tracking accuracy*, but this usage should not be taken to imply that mean-level bias is not related to the overall accuracy obtained in judgments.

In the example just given, positive mean-level bias is measured in terms of a tendency to overshoot or undershoot some benchmark. To measure accuracy, in whatever way, one or more benchmarks are always required. Much research uses within-individual benchmarks assessed cross-sectionally. For example, Boyes and Fletcher (2007, Study 3) had 57 heterosexual couples rate themselves and their partners on a range of scales, including judgments of attractiveness. Using 7-point scales, both men and women rated their partners (men: $M = 4.70$, women: $M = 4.77$) as significantly more attractive than they perceived themselves (men: $M = 4.14$, women: $M = 4.15$; $p < .01$)—a within-individual comparison. This is a classic example of the partner-serving bias found many times in the relationship literature.

However, a more objective benchmark consists of the self-judgments of the person being rated. Using this approach, Boyes and Fletcher (2007) also found that men and women exhibited a significant partner-serving bias, rating their partners (men: $M = 4.70$, women: $M = 4.77$) as more attractive than the self-perceptions of their partners (women: $M = 4.15$, men: $M = 4.14$)—an across-partner comparison. These results are especially striking given that the self-judgments of the partners used as benchmarks are themselves likely to be positively biased. A wealth of research and theory has documented the tendency for individuals to self-enhance routinely on such judgments (Sedikides, Gaertner, & Toguchi, 2003; Taylor & Brown, 1988).

In the current meta-analysis, we analyzed results only from studies that used objective benchmarks to assess mean-level bias (most commonly the self-judgments of the partner). However, we also included studies in which both judgment and benchmark came from the same person but that assessed benchmark judgments that were in the past (e.g., memories of previous levels of relationship satisfaction) or in the future (e.g., relationship dissolution). The general restriction of requiring an external benchmark, we think, produces more robust and theoretically interesting results than if we had included the plethora of studies that report within-individual judgments assessed concurrently. In addition, this means that those studies fitting our criterion (using external benchmarks), that happen to assess both mean-level bias and tracking accuracy, will use the same two variables (judgment vs. benchmark) in producing the two related effect sizes. Thus, using such studies, we were able to assess the extent to which the two ways of attaining accuracy are empirically independent in a robust fashion (given that the same variables and samples are involved in producing both indices).

Even a cursory appraisal of the empirical literature reveals that individuals in romantic relationships tend to evince quite substantial levels of tracking accuracy of partner judgments of many

kinds, ranging from personality attributions to mind readings of the partners' emotions and thoughts during interactions. The evidence regarding mean-level bias seems more variable, with findings of a relationship or partner-serving bias being quite common (as indicated previously), but with evidence of overall levels of negative mean-level bias at the sample level sometimes being produced as well (e.g., Archer, 1999; Friesen, Fletcher, & Overall, 2005). However, reviews in the relationship domain have been partial and quite restrictive in the research considered, so we do not know to what extent such findings generalize across different research domains.

In the meta-analysis to be described, we found that relevant research could be reliably categorized into six distinct categories: (a) judgments of partner personality traits; (b) judgments of negative attitudes, beliefs, or behaviors (e.g., aggression or criticism) directed to the partner or relationship; (c) judgments of positive attitudes, beliefs, or behaviors that are specifically directed to the partner or relationship, such as love or forgiveness; (d) mind-reading judgments (e.g., intentions and emotional states) advanced during interaction; (e) memories of past events or relationship states; and (f) predictions of future events or relationship states such as relationship satisfaction or longevity. The extent to which the effects are similar across such categories will attest to the robustness of the phenomena and may well raise further questions.

Categories b and c we refer to collectively as interaction traits. These traits are especially important in close relationships because they are specifically tied into a key feature of close relationships—namely, their interdependence (Kelley, 1979). Taking a leaf out of interdependence theory, Reis and colleagues (Reis & Patrick, 1996; Reis & Shaver, 1988) have argued that a key element in developing intimacy is the way in which the partner responds; specifically, to what extent does the partner communicate that he or she understands, validates, and cares for the other? The associated kinds of interaction attributions for how the individuals in a relationship think, feel, or behave toward each other (e.g., partner love or criticism) have turned out to be important in intimate relationships, consistent with the earlier theories of Kelley (1979) and Reis and colleagues (Reis & Patrick, 1996; Reis & Shaver, 1988; see also Murray & Holmes, 2009; Reis, Clark, & Holmes, 2004).

The empirical association between mean-level bias (positive or absolute) and tracking accuracy has, to our knowledge, not been investigated within specific studies in the relationship field. As stated previously, the current meta-analysis identified a subset of (38) studies that measured both mean-level bias and tracking accuracy using the same judgments and benchmark variables; thus, we were able to correlate the relevant effect sizes across studies. The outcome remains an open question. Nevertheless, we suspected that the correlation between the two ways of assessing accuracy may be slight, partly because of the prior example revealing that they are potentially distinct, but also because a preliminary analysis of the prior literature suggests that the two kinds of outcomes may be linked to two quasi-independent psychological systems each having different functions.

Before outlining our general model of relationship cognition, we turn briefly to two other kinds of biases in social judgment that are often discussed in the literature. To avoid confusion, it is important to distinguish these biases or errors from the two already described (mean-level bias and tracking accuracy), but they are also relevant to our model, some of the research discussed later, and to the general issue of the accuracy of lay judgments in relationships.

Clearing the Conceptual Undergrowth: Two More Kinds of Bias to Consider

Analyzing bias in terms of mean differences is a common method adopted in the research literature. However, bias is also often understood in the general (social) psychological literature, including the relationship domain, in terms of the undue influence that prior beliefs or lay theories are often considered to exert on specific judgments or perceptions (often termed belief perseverance). Indeed, the way in which prior beliefs or theories should be (and are) adjusted in the light of new evidence is at the heart of the study of rationality and bias in human judgment as well as in the philosophy of science (Fletcher, 1995). This kind of general bias is considered in detail in our proposed model, in those aspects of the meta-analysis that concern the moderating role of relationship quality, and later in the Discussion section.

Another judgmental process sometimes referred to as a bias (e.g., Kenny and Acitelli, 2001) concerns the extent to which individuals project the self when making partner judgments. Projection is sometimes viewed as an artifact and therefore as something to be statistically controlled for. For example, it is possible that higher levels of relationship satisfaction are associated with either tracking accuracy or mean-level bias because happier couples tend to project more and also happen to be more similar. On the other hand, projection can also be viewed as a heuristic that individuals use to attain accuracy. If couples are similar to varying degrees on certain traits or dimensions, then projecting appropriately on these traits should lead to more accuracy. Studies examining projection were not included in the meta-analysis, because they use complex path-analytic models, but they are reviewed later in the Discussion section in the light of the prior results and analysis and the model we next describe.

Explaining Positive Mean-Level Bias and Tracking Accuracy in Intimate Relationships

General Features of the Model

The general model underpinning our analysis is shown in Figure 1. The model is primarily intended as a general organizing framework that should be helpful in generating questions, suggesting directions for further research, solving puzzles, and integrating and explaining prior research findings. Because the model is pitched at a general level, it does not on its own generate many specific predictions. However, when combined with more specific theories, it is capable of generating (sometimes novel) predictions, which we illustrate in due course.

As can be seen in this model, it is assumed that the classic goals of social cognition (explanation, evaluation, prediction, and regulation, also often termed control) apply to relationship cognition. However, the additional goal of attaining relationship satisfaction is more or less unique to intimate relationships and is intended to capture a collection of aims such as finding a mate, developing a warm and satisfying sexual liaison, and so forth. Indeed, there is a massive literature showing that such goals are alive and well in intimate relationship settings (for reviews and edited collections

exemplifying this claim, see Baldwin, 2006; Fletcher & Fitness, 1996; Fletcher, Overall, & Friesen, 2006; Fletcher & Fincham, 1991).[4] These goals are also often linked, rather than stand alone, and are (like most social or person judgments) typically infused with affect. Accordingly, many relationship judgments are examples of hot rather than cold cognition.

The goals shown in Figure 1, in turn, motivate the formation of lay relationship theories that crucially come in two forms: general and local theories. General theories predate local or specific theories that begin their development as soon as two people meet for the first time. *General (relationship) lay theories* is an umbrella term including such dispositions as attachment working models, expectations, beliefs about the causes of relationship success, and mate value standards. These kinds of dispositions are distinct from individual differences that are not directly about intimate relationships, and are listed as a class of moderating variables (see Figure 1). These would include personality factors, self-esteem, cognitive complexity, and so forth. Local theories represent knowledge structures concerning a specific partner and associated relationship, and contain knowledge, beliefs, evaluations, and so forth, which are dispositional in nature.[5] Key components of this category comprise evaluations of the relationship (e.g., love, trust, commitment) and judgments of how the partner feels about the self and the relationship (interactive traits).

As can be seen in the model (see Figure 1), it is posited that goals influence lay theories and specific judgments, and vice versa. In Figure 1, mean-level bias and tracking accuracy are qualities of specific judgments rather than stand-alone categories (as shown by the use of double-headed arrows and dotted lines); indeed, they will often apply to the same judgments (as with our prior examples). The distinctions among the three categories shown in circles (general relationship theories, local relationship theories, and specific judgments) are fuzzy, given the overlap among these three categories. The category of specific judgments refers to online judgments that occur episodically, whereas the general and local theories are stored knowledge structures. Specific judgments about partners or relationships may be simply lifted from local theories and rendered conscious and accessible, for example, by being asked a question such as "How much does your partner love you?" Conversely, these items include judgments that are novel and situated in specific contexts, such as reading a partner's mind during an interaction or making an attribution for a specific behavior (Table 2 shows many examples of both kinds of specific judgments as assessed in the extensive research in this area).

The validity of the distinction between general and local relationship lay theories has considerable support. For example, in the adult attachment field, evidence supports the model formulated by Collins and Read (1994) in which local attachment working models for specific relationships are nested under global attachment working models, with influence working from top to bottom and vice versa (see, e.g., Mikulincer & Arad, 1999; Overall, Fletcher, & Friesen, 2003; Pierce & Lydon, 2001). Knee, Patrick, and Lonsbary's (2003) work has similarly shown that mental models of how relationships should function (framed in terms of beliefs in either romantic destiny or growth) influence attributions about specific relationship events, the strategies used to cope with conflict, and so forth.

An underlying premise in our model is that the lay perceiver is motivated to retain consistency between local relationship theories, general relationship theories, and specific judgments of the partner and relationship. Indeed, we would argue, along with many others, this feature is a fundamental feature of lay (social) cognition (see Fletcher, 1995). Thus, discrepancies among elements across the three categories of knowledge structure should drive changes in the system. This is hardly a new insight, being a common feature of classic models in social psychology including balance theory (Heider, 1958) and cognitive dissonance theory (Festinger, 1957).

In the relationship field, Thibaut and Kelley's (1959) seminal formulation of interdependence theory also used a version of balance theory, postulating that individuals' judgments of relationship satisfaction are derived by comparing perceived relationship benefits with two kinds of standards: the benefits viewed as deserved and the perceived availability of alternative partners. Expanding this pivotal idea from interdependence theory, Fletcher, Simpson, and their colleagues (Fletcher, Simpson, & Thomas, 2000; Fletcher, Simpson, Thomas, & Giles, 1999; Overall, Fletcher, & Simpson, 2006) have shown that lower consistency between ideal partner standards (residing in the general theories category) and associated specific partner perceptions (on important dimensions such as warmth, attractiveness, and status) drives lower levels of relationship satisfaction (Fletcher et al., 1999), motivates more partner regulation (Overall et al., 2006), and leads to higher levels of relationship dissolution (Fletcher et al., 2000). Critically, they have also shown that the mental comparisons between standards and perceptions have more predictive power than either the standards or partner perceptions taken independently (Overall et al., 2006).

In summary, the kind of consistency postulated here is exceptionally general, ranging from simple affective consistency to deductive reasoning, to the sort of complex cognitive models in which associative networks or mental models produce specific judgments (predictions, attributions, evaluations, etc.). Turning to the more novel part of our model, we postulated that this cognitive system influences the levels of mean-level bias and tracking accuracy (see Figure 1). Moreover, the moderating factors are proposed to prime different aims, and thus alter the levels of mean-level bias and tracking accuracy.

---

[4] Nested under these general goals will exist a myriad of midlevel goals and specific goals. For example, an influential body of research investigates how people manage the psychological risks in the development of dependency and intimacy in romantic relationships. Work from an adult attachment perspective (Collins & Allard, 2001; Gillath et al., 2006), a risk regulation perspective (Murray, Holmes, & Collins, 2006), and a promotion versus avoidance motivation approach (Gable & Poore, 2008) all focus on two kinds of preferred goals: the goal of avoiding dependency, threats to the self, or relationship intimacy versus the goal of seeking and gaining the benefits of a secure, close relationship. These two kinds of goals are linked most directly to the more general goal in our model of achieving relationship satisfaction (see Strachman & Gable, 2006).

[5] We use the term *lay theories*, as opposed to *knowledge*, *structures prototypes*, or *beliefs*, because it captures various important characteristics of such knowledge structures, including (a) the way in which the items involved form relatively coherent mental packages or stories, (b) the nature of some of the associated functions such as explanation and prediction, and (c) the different levels at which such lay theory is constituted. Of course, there are also substantial differences between lay theories and scientific theories (see Fletcher, 1995, 2002).

Table 2
Summary Information From Studies Used in Meta-Analyses

Study	Country	Sample size	Relationship status	Gender[a]	Relationship length (years)	Sample population	Construct measured	Judgment category[b]	Category effect size[c]
Acitelli et al. (2001)	United States	238	Married	M & F	9.81	Community	Marital ideals	3	1, 3
Agnew et al. (2001)	United States	74	Dating	M & F	1.57	College	Breakups	6	1
Andrews et al. (2008)	United States	203	Dating	M & F	1.72	Community	Sexual infidelity	2	1, 2
Barelds-Dijkstra & Barelds (2008)	Netherlands	93	Married	M & F	13.50	Community	Attractiveness	1	2
Bell et al. (2008)	United States	244	Married	F	6.35	Clinical	Aggression	6	1, 2
Boyes & Fletcher (2007)	New Zealand	57	Dating	M & F	2.27	College	Personality traits and attributes	1	1, 2, 4
Calhoun et al. (2002)	United States	94	Married	F	—	Clinical	Anger trait	1	1, 2
Capaldi & Owen (2001)	United States	318	Married	M & F	1.71	Clinical	Aggression	2	1
Capaldi et al. (2003)	United States	105	Dating	M & F	1.42	Clinical	Aggression	2	1
Cattaneo et al. (2007)	United States	246	Married	F	6.45	Clinical	Risk aggression	6	1, 2
Cattaneo & Goodman (2003)	United States	169	Married	F	2.10	Clinical	Risk aggression	6	1
Clements et al. (2007)	United States	19	Married	M & F	8.20	Community	Online emotions and cognitions	4	1, 3
Collins & Feeny (2000)	United States	93	Dating	M & F	1.05	College	Support behavior	3	1
Collins & Feeney (2004): Study 2	United States	153	Dating	M & F	1.24	College	Support behavior	3	1
Corsini (1956)	United States	20	Married	M & F	8.50	College	Personality traits	1	1, 3
Cui et al. (2005)	United States	472[d]	Dating	M & F	1.72	College	Hostile behavior	2	1, 2
De La Ronde & Swann (1998)	United States	81	Married	M & F	5.60	Community	Personality traits and attributes	1	1
Drigotas et al. (1999): Study 1	United States	53	Dating	M & F	1.76	College	Ideal self movement	1	1
Drigotas et al. (1999): Study 3	United States	50	Dating	M & F	1.13	College	Personality traits and self ideals	1	1
Dunn et al. (2007): Study 1	United States	56	Dating	M & F	1.09	College	Affect after interactions	6	2
Dunn et al. (2007): Study 2	United States	38	Dating	M & F	1.08	College	Affect after interactions	6	2
Eastwick et al. (2008)	United States	26[d]	Dating	M & F	0.26	College	Distress after breakups	6	1, 2, 4
Edwards & Klockars (1981)	United States	89	Married	M & F	7.62	Community	Personality traits	1	1
Epley & Dunning (2006): Study 3	United States	36	Dating	M & F	—	College	Breakups	6	1, 2
Foltz et al. (1999)	United States	49	Married	M & F	3.10	College	Interpersonal traits	1	1, 2
Försterling et al. (2005)	Germany	72	Dating	M & F	3.10	College	Attributions	6	1
Friesen et al. (2005)	New Zealand	39	Married	M & F	3.15	College	Forgiveness	3	1, 2, 4
Gagné & Lydon (2001): Study 2	Canada	108	Dating	M & F	1.37	College	Breakups	6	1
Gagné & Lydon (2001): Study 3	Canada	30	Dating	M & F	1.57	College	Breakups	6	1
Gagné & Lydon (2003): Study 1	Canada	47	Dating	M & F	1.56	College	Personality traits	1	1, 2, 4
Gagné & Lydon (2003): Study 2	Canada	105	Dating	M & F	2.19	College	Personality traits and attributes	1	1, 2, 4
Gauthier et al. (2008)	Canada	58[d]	Married	M & F	10.6	Community	Pain estimation	4	1
Gomà-i-Freixanet (1997)	Spain	198	Married	M & F	10.00	Community	Personality traits	1	1, 2
Goodwin et al. (2002): Studies 1–2	United States	122[d]	Single (dating context)	M & F	N/A	College	Social appeal trait	1	2
Goodwin et al. (2002): Studies 3–4	United States	112[d]	Single (dating context)	M & F	N/A	College	Personality traits and attributes	1	2
Heckert & Gondolf (2004)	United States	348	Married	F	—	Clinical	Risk aggression	6	1
Holzworth-Munroe et al. (2000)	United States	102	Married	F	9.50	Community	Hostility	2	2
Julian et al. (1999)	United States	152	Married	F	13.11	Community	Aggression	2	1

(*table continues*)

Table 2 (*continued*)

Study	Country	Sample size	Relationship status	Gender[a]	Relationship length (years)	Sample population	Construct measured	Judgment category[b]	Category effect size[c]
Kahn (1970)	United States	42	Married	M & F	5.17	College	Online nonverbal affect	4	1, 3
Karney & Coombs (2000)	United States	91	Married	F	14.1	College	Relationship quality	5	1, 2, 4
Karney & Frye (2002): Study 1	United States	74[d]	Married	M & F	6.90	Community	Relationship quality	5	1, 2
Karney & Frye (2002): Study 2	United States	237[d]	Married	M & F	6.75	Community	Relationship quality	5	1, 2
Kenny & Acitelli (2001)	United States	213	Married	M&F	10.00	Community	Positive attitudes and interaction	3	1
Khawaja & Tewtel-Salem (2004)	Lebanon	417	Married	M & F	14.5	Community	Aggression	2	1
Kilpatrick et al. (2002)	United States	55	Married	M & F	5.30	Community	Online emotions and cognitions	4	1, 3
Kirchler (1999)	Austria	286	Married	M & F	19.36	Community	Conflict tactics	2	1, 3
Koerner & Fitzpatrick (2002)	United States	64	Married	M & F	15.20	Community	Online nonverbal affect	4	1
Kunda (1987): Study 1	United States	103	—	—	—	College	Attributions	1	2
Lawrence & Bradbury (2001)	United States	56	Married	M & F	2.75	Community	Aggression	2	1, 2
Lawrence & Bradbury (2007)	United States	164	Married	M & F	3.25	Community	Aggression	2	1
Lemay & Clark (2008) study 4	United States	102	Married	M & F	3.54	Community	Responsiveness to needs	3	1, 3
Lemay et al. (2006)	United States	290	Married	M & F	43.9	Clinical	Illness treatment preferences	4	1, 2, 3, 4
Lorenz et al. (2007)	United States	292[d]	Married	M & F	—	Community	Hostility	2	1, 2
Loving (2006)	United States	284/ 108[d]	Dating	M & F	0.47	College	Breakups	6	1, 2
MacDonald & Ross (1999)	Canada	76[d]	Dating	M & F	0.50	College	Breakups	6	1
Marshall & Holtzworth-Munroe (2002)	United States	164	Married	F	9.47	Community	Aggression	2	1, 2
McCrae (1982)	United States	281[d]	Married	M & F	—	Community	Personality traits	1	1
McCrae (1993)	United States	68[d]	Married	M & F	—	Community	Personality traits	1	1
McCrae et al. (1998)	United States	47	Married	M & F	38.00	Community	Personality traits	1	1, 3
McFarland & Ross (1987)	Canada	68[d]	Dating	M & F	1.09	College	Personality traits	1	1
Murray, Bellavia, et al. (2003)	United States	154	Married	M & F	10.10	Community	Love	3	1
Murray et al. (2002): Dating	United States	86	Dating	M & F	2.00	College	Personality traits, values, and feelings	1	1, 3
Murray et al. (2002): Married	Canada	93	Married	M & F	10.90	Community	Personality traits, values, and feelings	1	1, 3
Murray et al. (1996a): Dating	Canada	98	Dating	M & F	1.58	College	Personality traits	1	1, 2, 3, 4
Murray et al. (1996a): Married	Canada	75	Married	M & F	6.50	Community	Personality traits	1	1, 2, 3, 4
Murray et al. (1996b)	Canada	84	Dating	M & F	1.58	College	Personality traits	1	1, 2, 3, 4
Murray et al. (2001)	Canada	105	Married	M & F	10.9	Community	Love	3	1
Mutén (1991)	United States	72[d]	Married	M & F	—	Clinical	Personality traits	1	1
Noller (1980)	Australia	48	Married	M & F	12.56	Community	Online nonverbal affect	4	1, 2, 3, 4
Noller & Feeny (1994)	Australia	33	Married	M & F	3.66	Community	Online nonverbal affect	4	1
Oishi & Sullivan (2006)	United States	79	Dating	M & F	0.17	College	Behaviors, relationship quality	5	1, 2
O'Leary & Williams (2006)	United States	453	Married	M & F	9.50	Community	Aggression	2	1
Overall et al. (2006)	New Zealand	62	Married	M & F	2.82	College	Personality traits and attributes, ideals, regulation	2	1, 2

Table 2 (*continued*)

Study	Country	Sample size	Relationship status	Gender[a]	Relationship length (years)	Sample population	Construct measured	Judgment category[b]	Category effect size[c]
Panuzio et al. (2006)	United States	303	Married	M & F	13.2	Clinical	Aggression	2	1, 2
Perry & Fromuth (2005)	United States	50	Dating	M & F	0.85	College	Aggression	2	1
Rhoades & Stocker (2006)	United States	119	Married	M & F	21.93	Community	Aggression	2	1, 2
Ro & Lawrence (2007)	United States	100	Married	M & F	3.09	Community	Hostility	2	1
Robinson & Price (1980)	United States	56	Married	M & F	—	Community	Positive behavior	3	1, 4
Sabatelli et al. (1982)	United States	48	Married	M & F	5.54	Community	Online nonverbal affect	4	1
Saffrey et al. (2003)	Canada	76	Married	M & F	4.01	College	Interpersonal traits	1	2, 3
Sagrestano et al. (1999)	United States	42	Married	M & F	7.18	Community	Aggression	2	2
Sanderson & Cantor (2001)	United States	44	Married	M & F	18.85	Community	Intimacy goals	3	1, 2
Sanderson & Evans (2001)	United States	100	Dating	F	1.25	College	Intimacy goals	3	1, 2
Schul & Vinokur (2000): Study 1	United States	627[d]	Married	M & F	—	Community	Depression	1	1, 2
Schul & Vinokur (2000): Study 2	United States	227[d]	Married	M & F	—	Clinical	Traits, attributes, events	1	1
Senécal et al. (2003)	Canada	36	Married	M & F	2.6	College	Emotions	6	2
Sillars et al. (1984): Study 1	United States	33	Married	M & F	10.20	Community	Communication conflicts	2	1, 3
Sillars et al. (1984): Study 2	United States	40	Married	M & F	10.80	Community	Communication conflicts	2	1, 3
Sillars et al. (1990)	United States	37	Married	M & F	11.10	Community	Negative communication	2	3
J. A. Simpson et al. (2003)	United States	95	Married	M & F	7.79	Community	Online emotions and cognitions	4	1, 3
L. E. Simpson & Christensen (2005)	United States	273	Married	M & F	11.90	Community	Aggression	2	1, 2
D. A. Smith & Peterson (2008)	United States	72[d]	Married	M & F	12.22	Community	Criticism	2	1, 2, 4
L. Smith et al. (2008)	Australia	81	Married	M & F	1.23	Community	Emotional intelligence	3	1, 2
South et al. (2004)	United States	3,096	Married	M & F	—	Community	Divorce	6	1
Sprecher (1999)	United States	59	Married	M & F	4.06	College	Relationship quality	5	1
Swann & Gill (1997)	United States	57	Dating	M & F	1.50	College	Personality traits and attributes and events	1	1, 3
Thomas & Fletcher (2003)	New Zealand	50	Dating	M & F	1.38	College	Online emotions and cognitions	4	1, 3
Thomas et al. (1997)	New Zealand	74	Married	M & F	17.70	College	Online emotions and cognitions	4	1, 3
Tiggle et al. (1982): Study 1	United States	77	Married	M & F	11.50	College	Desire for partner change	2	3
Tiggle et al. (1982): Study 2	United States	75	Married	M & F	5.00	College	Desire for partner change	2	3
Tucker & Anders (1999)	United States	61	Dating	M & F	1.00	College	Relationship evaluations	3	1, 3
Van Lange et al. (1997): Study 4	United States	45	Dating	M & F	1.58	College	Willingness to sacrifice	3	1
Van Lange et al. (1997): Study 6	United States	64	Married	M & F	4.83	Community	Willingness to sacrifice	3	1
Waldinger & Schulz (2006)	United States	152	Married	M & F	7.10	Community	Online intentions and attributions	4	1
Watson et al. (2000): Dating	United States	272	Dating	M & F	1.52	College	Personality and affective traits	1	1
Watson et al. (2000): Married	United States	74	Married	M & F	19.38	Community	Personality and affective traits	1	1
Watson & Humrichouse (2006)	United States	301[d]	Married	M & F	4.00	Community	Personality amd affective traits	1	1, 2
Weigel (2008)	United States	121	Married	M & F	9.10	College	Commitment behaviors	3	1, 2, 4

(*table continues*)

Table 2 (*continued*)

Study	Country	Sample size	Relationship status	Gender[a]	Relationship length (years)	Sample population	Construct measured	Judgment category[b]	Category effect size[c]
Weisz et al. (2000)	United States	177	Married	F	—	Clinical	Risk aggression	6	1
Werner et al. (2001)	United States	202	Married	M & F	15.20	Community	Negative communication	2	1
Wilhelm & Perrez (2004)	Switzerland	95	Married	M & F	21.87	Community	Daily-life feelings	4	1, 2, 3

Note. Sample size represents the number of couples, except where gender indicates the sample was composed of men or women. For Loving (2006), the sample size is 284 for tracking accuracy and 108 for mean-level bias. Dashes indicate that information was unavailable. N/A = not applicable.
[a] M = male, F = female. [b] 1 = personality traits, 2 = negative interaction traits, 3 = positive interaction traits, 4 = mind reading, 5 = memories, 6 = predictions. [c] 1 = tracking accuracy, 2 = mean-level bias, 3 = association between tracking accuracy and relationship evaluation, 4 = association between positive mean-level bias and relationship evaluations. [d] Sample includes men and women from different relationships.

For example, the model would predict that when the goal of relationship satisfaction or evaluation is primed, this should motivate those who place higher value on physical attractiveness to exaggerate their partners' attractiveness (but not influence tracking accuracy). In contrast, when the more epistemic goals of prediction or regulation are primed, this should lead to increases in tracking this trait accurately (but not to changes in mean-level bias). However, when a trait that is less relevant in the context of sexual relationships is being judged (such as talkative or artistic), neither positive mean-level bias nor tracking accuracy should change (for detailed predictions of how truth-seeking vs. positivity-seeking motives should influence positive mean-level bias and tracking accuracy, see Fletcher, Simpson, & Boyes, 2006).

The model (see Figure 1) also predicts that changes in the specific judgments in terms of mean-level bias and/or tracking accuracy can influence stored relationship theories, which may, in turn, lead to changes in the accessibility or nature of the goals.[6] For example, we would predict that if levels of positive mean-level bias and tracking accuracy were enhanced, on important partner traits (such as honesty, sensitivity, and warmth), this would lead people to be more positive about their relationships and feel closer to their partners. Indeed, this is precisely what Lackenbauer, Campbell, Fletcher, and Rubin (in press) have found. Thus, overall, levels of bias and tracking accuracy are a product of the drive for consistency among the components in the model and are also key elements in allowing consistency to be maintained.

In summary, the love-is-blind hypothesis posits that judgments of partners and relationship judgments are primarily motivated by wishful thinking designed to stamp out doubt and maintain high levels of commitment and judgments of relationship quality (see Fletcher, 2002; Fletcher & Boyes, 2008). Indeed, one of the apparent demonstrations of this postulate often appealed to (as previously noted) is the empirical finding that more positive mean-level bias (using objective benchmarks) is associated with higher relationship satisfaction. We expected to confirm this finding from the results of the current meta-analysis.

If it is the case that relationship satisfaction remained immune from observations or evidence to the contrary, then this finding would support an extreme version of the love-is-blind hypothesis. However, there is abundant evidence showing that relationship satisfaction changes over time as a function of the quality of objectively assessed communication (Fletcher & Thomas, 2000), perceptions of problem severity in the relationship (Overall, Fletcher, Simpson, & Sibley, 2009), positive expectations (McNulty & Karney, 2004), and how positively individuals believe their partners view them (Murray, Griffin, Rose, & Bellavia, 2003). This degree of malleability is inconsistent with a strong version of the love-is-blind hypothesis.

Applying this model to existing research can also change the standard interpretations of such findings. Take, as an example, a commonly cited article by Murray and Holmes (1993) in which participants were persuaded to think (by reading bogus *Psychology Today* articles) that either open disagreement (Study 1) or recognizing differences between partners (Study 2) is crucial to the development of intimacy in romantic relationships. Prior to these experimental manipulations, participants completed short exercises designed to buttress perceptions of their partners as rarely initiating disagreement or as being very similar to the self, respectively. The results convincingly showed how participants (relative to controls) resolved their freshly minted doubts either by embellishing the presence of open disagreement or couple differences or by emphasizing or reinterpreting other aspects of their relationship to counter the suddenly apparent weakness. In short, the participants seemed to crush doubt and maintain their rose-tinted views by simply rewriting history and altering some key relationship perceptions—a process that looks at face value as anything but scientific or rational.

Note first that this study cleverly engineered a loss of fit between local relationship theories and general theories. The highlighted results showed that participants commonly responded by altering some detailed components of their local relationship accounts. However, other results (that were not the primary focus) suggested that some participants developed more negative relationship evaluations, that some rejected the experimenter-supplied general theory completely, and that the manipulation (not surprisingly) did not work for participants who possessed prior local and relationship theories that were close to the manipulated versions. In terms of the model in Figure 1, the participants overall seemed to have been principally motivated to maintain consistency be-

[6] The distinction between controlled and automatic cognition is also an important element in models of relationship cognition (see, e.g., Fletcher, 2002; Murray & Holmes, 2009). However, we have omitted this distinction in this model because of space limitations and because very little research in this domain is linked to this distinction.

tween their local theories, their general theories, and their relationship judgments. In addition, the findings illustrate a key point made previously that changes in specific judgments are both a product of strains between general and local theories and a key mechanism in allowing consistency to be retained.

Models such as the one we propose explain why and how lay (relationship) theories (just like scientific theories) are typically slow to change in the face of inconsistent data: The system is conservative, and for good reason, given the momentous consequences of sudden changes in levels of commitment or models of the partner and relationship. At the most general level, if the stored theories are sounder and more informed, this should lead to higher tracking accuracy (all being equal). If the theories are more loaded with positive affect, this should lead to more positive mean-level bias in specific judgments (all being equal).

However, the real explanatory power of the model comes into play when explaining what happens when the system is put under stress by introducing a potential or actual inconsistency. In this case, the results for mean-level bias or tracking accuracy will be a function of how consistency among the elements is retained, and this will differ as a function of other moderating variables and the associated goals that are primed.

For example, experimental research on attachment has shown that when relationship-threatening events are introduced, such as poor partner support, individuals with anxious relationship working models produce judgments of their partner that are more negative in terms of mean-level bias (e.g., Collins & Feeney, 2004) but higher in tracking accuracy (e.g., J. A. Simpson, Ickes, & Grich, 1999). They are also more likely to develop more negative relationship satisfaction over time (J. A. Simpson et al., 1999).

In terms of our model, when responding to lack of partner support (a moderator in Figure 1), anxious attachment working models (a species of general relationship theory that involves strong approach–avoidance tendencies) promote a negative spin on partner judgments, but also higher levels of tracking accuracy as a function of increased vigilance and the salience of the goal of prediction (e.g., can I trust my partner?). We give many other examples in the Discussion section of how relevant research findings fit into the proposed model.

In short, when the equilibrium of the cognitive system is disturbed, specific judgments may or may not change, depending on other moderating variables and the goals. Sometimes, mean-level bias or tracking accuracy may actually stay the same, and local relationship theories may shift to satisfy the need for cognitive consistency. This point is exemplified in our discussion of the prior research by Murray and Holmes (1993) in which a loss of fit between local relationship theories and general theories led some people either to alter detailed components of their local relationship accounts or to change their local relationship theory.

Predicting which strategy individuals will adopt to produce consistency among the major components in the model is not an easy matter. A Bayesian approach offers some guidelines here; namely, relevant local or general theories that are held strongly should be likely to produce changes in local judgments (including mean-level bias and tracking accuracy). In contrast, local or general theories that are weakly held should be more likely to change when new information challenges the system. However, although our model may also prove helpful in this regard, the ability to predict exactly when, which, and how people resolve such cognitive inconsistencies remains a work in progress.

Of the list included in Figure 1, only two moderating factors have been researched sufficiently to warrant inclusion in the following meta-analyses: relationship stage and gender. Some of the other moderating factors listed have also been investigated occasionally, but these are discussed later in the article. However, from here on we confine the material discussed in the introduction to these two moderators, along with the role played by relationship evaluations (or satisfaction), which in our model is represented as a central component of stored local relationship theories.

Mean-Level Bias

As noted, we assessed the amount of mean-level bias evinced at the sample level across studies. Studies that report evidence of an overall positive partner-serving or relationship-serving mean-level bias are commonplace, as exemplified by the Boyes and Fletcher (2007) research described previously. However, some research has reported the existence of overall negative mean-level bias. For example, Friesen et al. (2005) found that individuals perceived their partners as forgiving them less for specific infractions than their partners reported. In a meta-analysis of studies using the Conflict Tactics Scales (Archer, 1999), individuals perceived their partners as being more violent toward them than the partners' reports suggested (a negative partner bias).

Of course, such differences in overall positive mean-level bias across samples could simply reflect sample-level differences in relationship quality. However, error management theory (Haselton & Buss, 2000) suggests an explanation in terms of the primacy of the goal of maintaining a successful relationship; namely, the relationship costs involved in holding prior positive versus negative biases, which may differ according to the kind of judgment being considered. In the example used by Haselton and Buss (2000), an effective smoke fire alarm is designed to be sensitive to small amounts of smoke and heat, which produces many false alarms. In this context, a false negative (not going off when a fire is present) is much more harmful than a false positive (a false alarm). When the judgment involved consists of an interaction trait that is focused on the connection between the self and the partner (such as relationship violence, or forgiveness, or love), it seems plausible that the default bias might be set on the negative side. If one overestimates the forgiveness, trust, or love of a partner for the self, this might lead to complacency and lack of effort in building a more secure relationship. In contrast, positive biases such as perceiving one's partner as more attractive, kind, and ambitious than he or she really is seem to have fewer obvious downsides. Such individual-level attributions refer to general traits that may have strong implications in relationships but nevertheless are not focused specifically on dyadic interaction.

In terms of our model in Figure 1, this explanation exploits the importance of the goals of maintaining a successful relationship and appropriately regulating self and partner behavior. It also suggests that the kind of specific judgments produced makes a difference in terms of mean-level bias. In summary, we expected that the mean sample-level amounts of positive bias should vary across judgment domains in the fashion outlined, with low or even negative amounts of mean-level bias produced when interaction attitudes or traits are being judged.

Tracking Accuracy

What is the role of tracking accuracy in romantic relationships? One possibility is that it serves to limit the costs inflicted by donning rose-tinted glasses, given the additional understanding conferred by understanding the overall structure of the target's personality and accurately ascertaining his or her relative strengths and weaknesses. Romantic relationships also afford both the knowledge and the motivation to understand the beliefs, traits, and attitudes of the partner. Thus, we expected that samples of individuals would reveal substantial levels of tracking accuracy across all classes of judgment assessed including partner personality traits, positive or negative interaction traits, mind-reading judgments during interaction, and memories and predictions of relationship states or events.

In addition, there is already evidence for a basic prediction from our model, previously stated, that sounder and more informed relationship theories should lead to higher tracking accuracy. For example, tracking accuracy tends to increase with more information and longer acquaintance (Letzring, Wells, & Funder, 2006), and is higher in married and dating samples than among friends or strangers (Beer & Watson, 2008; Funder & Colvin, 1988; Watson, Hubbard, & Wiese, 2000).

As a specific research example, Thomas and Fletcher (2003) had couples in romantic relationships produce mind readings of one another, based on 10-min problem-solving discussions, using the standard video review procedure developed by Ickes, Stinton, Bissonnette, and Garcia (1990). Friends of one of the partners and strangers also viewed the same tapes and produced mind readings of the participants' thoughts and feelings. Results showed that dating partners generally achieved the highest mind-reading accuracy followed by their friends and in turn the strangers. Moreover, other analyses suggested that the advantage enjoyed by insiders in the relationship was a function of the additional knowledge and more insightful local theories they possessed in contrast to the relationship outsiders.

Relationship Satisfaction and Moderators of Positive Mean-Level Bias

If the meta-analysis reveals a substantial link between positive mean-level bias (using objective benchmarks) and relationship satisfaction, then this would support the claim that mean-level bias serves to buttress the kind of love and commitment that is associated with long-term bonding in humans. It also serves to support a key postulate in the model (see Figure 1) that the drive for cognitive and affective consistency produces a positive correlation between relationship evaluations (part of local relationship theories) and the positive mean-level bias inherent in specific judgments.

If such a positive bias, in turn, is driven by the kind of strong emotions associated with love at its peak, then the strength of this association should be linked to the stage of the relationship (a moderating variable shown in Figure 1). Passionate love (linked to earlier stages of the relationship) should be strongly linked to positive mean-level bias, whereas companionate love—which is no less committed but has lost its urgency and fire, being linked to later stages of the relationship—should be weakly linked to positive mean-level bias. In the meta-analysis, we assessed the moderating influence of relationship length. Based on this reasoning, a clear-cut prediction is that as the mean time that samples have been in their relationships increases across studies, we should find weaker associations between relationship satisfaction and amounts of positive mean-level bias (as indexed by the associated effect sizes).

Some studies have reported evidence that women are more prone to positive mean-level bias than men, especially in dating relationships (see Gagné & Lydon, 2003), suggesting that gender may also be an important moderating variable (see Figure 1). Gagné and Lydon's (2003) explanation for this finding exploits the idea that women are more oriented toward the maintenance of successful intimate relationships than men and therefore slip into partner-serving biases more readily than men in the beginning stages of romantic relationships. In contrast, men may be more likely to view budding relationships in casual, short-term, or utilitarian terms and thus require higher levels of love and commitment than women before donning rose-colored glasses.

This explanation is consistent with parental investment theory (Trivers, 1972), which proposes that the sex that invests more in the offspring will be choosier in terms of mate selection. It is also consistent with a raft of findings showing that women are more focused than men on relationship maintenance, more likely to be the relationship managers than men, and choosier in selecting mates than men on dimensions linked to investment such as status and resources (for a review, see Fletcher, 2002).

Gagné & Lydon (2003) tested this explanation by showing that commitment to the relationship was more strongly associated with higher levels of such positive biases for men than for women (although the evidence was mixed across two studies). We were able to test Gagné and Lydon's explanation further by assessing gender differences in the extent to which amounts of mean-level bias were associated with relationship evaluations.

If we combine error management theory with the ideas put forward by Gagné and Lydon (2003), then this implies that gender (a moderating variable in Figure 1) might interact with the kind of specific judgments produced to influence mean-level bias. As we have already argued, for general attributions focused at the individual level (e.g., partner attractiveness), the relationship costs for negative bias are probably higher than for positive bias. In contrast, for attributions of interaction traits that are located within specific relationships (such as forgiveness), this pattern should reverse, with the costs for positive bias being higher than for negative bias. If women are more inclined to invest heavily in romantic relationships than men, then women should more closely monitor relative, potential relationship rewards and costs.

Thus, women should be more prone to positive bias than men for individual-level attributions (as has been shown) but more prone to negative bias for interaction traits. We were able to test this ambitious, novel prediction using selected studies from the meta-analysis.

Relationship Satisfaction and Moderators of Tracking Accuracy

A favorite prediction of researchers has been that higher accuracy should be associated with more relationship satisfaction. However, the research findings with respect to tracking accuracy reveal a mixed picture, with some studies reporting positive cor-

relations (e.g., Kilpatrick, Bissonnette, & Rusbult, 2002), some reporting null findings (Thomas & Fletcher, 2003), and yet others citing negative correlations between tracking accuracy and relationship satisfaction (Sillars, Pike, Jones, & Murphy, 1984). However, no prior systematic meta-analysis has been published of the published findings. If there turns out to be little overall evidence across studies of a positive link between relationship satisfaction and tracking accuracy, this would suggest that tracking accuracy is not hooked into the evaluative relationship system in the same way that positive mean-level bias seems to be.

We suggested above that relationship length should moderate the link between relationship satisfaction and positive mean-level bias (see Figure 1). Along similar lines, Thomas and Fletcher (2003) found that relationship length moderated the association between relationship satisfaction and tracking accuracy; that is, relationship satisfaction was found to be higher in shorter relationships to the extent that partners were inaccurate mind readers (love perhaps motivates inaccuracy), whereas relationship satisfaction was enhanced in longer relationships as tracking accuracy increased (for a similar finding, see Campbell, Lackenbauer, & Muise, 2006).

The explanation is similar to that offered previously for mean-level bias; namely, passionate love (linked to earlier stages of the relationship) might focus attention on the need to maintain relationship satisfaction, whereas companionate love (linked to later stages of the relationship) should encourage and permit deeper partner knowledge to be garnered and exploited with lower levels of attendant threat. Thus, we should find that as relationship longevity increases across samples, the effect sizes between tracking accuracy and relationship satisfaction will also tend to increase (analogous to our prediction for mean-level bias).

Consistent with the arguments made previously, there is good evidence that women are generally more motivated and expert lay psychologists than men (for a review, see Fletcher, 2002). Thus, women may be motivated to develop more accurate partner and relationship judgments than men. However, the findings are mixed in terms of tracking accuracy, with some studies finding clear evidence of the superiority of women (e.g., Thomas & Fletcher, 2003) and others finding no evidence of gender differences (e.g., Acitelli, Kenny, & Weiner, 2001). In the current meta-analysis, we were able to test for the existence of a reliable gender difference in tracking accuracy across several judgment domains.

The Current Meta-Analysis

In the current meta-analysis, we reviewed an extensive set of studies that had examined mean-level bias (48 studies) and tracking accuracy (98 studies) of judgments in intimate sexual relationships, using some kind of objective criterion against which to compare the judgments. The aims were to test the reliability and generality of effects that have been widely reported in the literature, to test our postulation that there exist two basic and independent forms of accuracy that characterize judgments of partners or relationships, and to test some explanations for these effects in terms of analyzing some moderating variables (length of relationship and gender). The results may help resolve the paradox that love appears to be both blind and rooted in reality.

The meta-analysis assessed studies that reported simple cross-sectional results or contained bivariate analyses that could be converted to standard effect sizes. More complex multivariate studies, path-analytic studies, longitudinal analyses, and experimental studies were also reviewed, but these are not readily handled in a meta-analysis. They are, however, discussed where relevant in the Discussion section, within the framework of the model in Figure 1, which we note goes well beyond the moderators and findings explicitly dealt with in the meta-analysis.

On the basis of our model, and some additional ideas and theories drawn from evolutionary psychology, social psychology, and prior research, we tested the following predictions in the meta-analysis: (a) significant levels of both positive and absolute mean-level bias should be found, with the exception of the category of interaction traits, which should reveal negative levels of signed mean-level bias; (b) significant levels of tracking accuracy should generally be revealed; (c) higher levels of positive mean-level bias should be associated with higher levels of relationship quality; (d) the links between mean-level bias and tracking accuracy with relationship satisfaction should decrease and increase, respectively, as a function of higher relationship length; (e) women should produce higher levels of tracking accuracy than men; and (f) women should produce higher levels of positive mean-level bias than men for individual-level attributions but more negative levels of mean-level bias for interaction traits.

Questions, with no predictions attached, included (a) to what extent do the findings generally vary across judgment categories? (b) is tracking accuracy linked with relationship quality? and (c) to what extent are mean-level bias and tracking accuracy independent in those studies that have assessed both aspects using the same variables and samples?

Method

Selection of Studies

Research that was published up to the end of 2008 was included in this meta-analysis. The body of research we were interested in was diverse and published in a wide array of journals. Thus, we used a large number of keywords in pairs for the searches. Each word from this initial list was conjoined in turn with each word from the second list in the search engines used (Web of Science and PsycINFO): List 1: *bias*, *accuracy*, *illusions*, *mind reading*, *understanding* (with *bias* only), *agreement*, *similarity*, *predictions*, *projection*; List 2: *romantic*, *partners*, *marriage*, *dating*, *marital*, *couples*, *intimate*, *relationships*, *love*, *spouses*, *dyadic*, *aggression*, *violence*, *traits*. We limited the references to psychology, social science, and relationship journals. We also read the reference lists of all the selected articles and, using Web of Science, traced the articles that had cited classic or commonly cited articles in the field. Because of the polyglot nature of the research coverage, and the wide time span involved, we examined only published, peer-reviewed research articles (not unpublished studies). Because there existed a previously published meta-analytic review of the research on accuracy and bias across partners for aggression in relationships by Archer (1999), we examined only studies in this particular domain from 1998 onward (i.e., articles not included in Archer's analysis).

Studies were included only if (a) they included samples of adult individuals (18 years or older) who were in heterosexual or homosexual relationships (platonic relationships were excluded) and

(b) the benchmark for bias or accuracy was external to the participant (i.e., not provided by the participant), with the exception noted previously of studies that involved memories of prior states of the relationship (e.g., relationship satisfaction) or predictions (e.g., relationship dissolution). Behavioral studies that included estimates of the bias or accuracy with which partners judged relationship or partner behavior were included. However, to qualify for the meta-analysis, the measures of the behavior and the judgment had to be equivalent. For example, a study that asked individuals to judge the warmth and support of the partner during problem-solving interactions and correlated the responses with observer ratings of the same behavior would qualify. However, if the study in question had correlated such observer ratings with self-reports of the stress produced, or the general supportiveness of the partner, or attributions for the behavior, then such findings were not included. This necessary criterion ruled out many studies.

Authors were contacted for additional information, where feasible, if it was required to assess the effect sizes. Ninety-eight studies were selected that included information on tracking accuracy and 48 that included information on mean-level bias. Of these, 38 studies were identified that included reports of both mean-level and tracking accuracy for the same samples and used the same variables. Studies that used difference scores to measure tracking accuracy were not included, given the well-known problems with this approach (Griffin, Murray, & Gonzalez, 1999). Finally, we identified 14 articles that included usable information on the link between mean-level bias and relationship satisfaction, and 27 studies that reported usable information on the link between tracking accuracy and relationship satisfaction (three of this latter group of studies did not report mean levels of tracking accuracy and so were not included in the meta-analysis of tracking accuracy per se; see Table 2).

Coding

Information gathered from the articles included (a) the authors, (b) date of publication, (c) country of origin, (d) sample size (actual sample size used in analyses reported), (e) relationship status (married or dating), (f) couples or individuals in sample, (g) gender of sample, (h) relationship length, (i) nature of the sample (clinical, community-based, or comprising university students), (j) nature of the construct measured, (k) judgment category, and (l) whether the study reported tracking accuracy, mean-level bias, and the links between either measure and relationship evaluations. We were open to studies including homosexual samples, but all studies located reported using samples of participants in heterosexual relationships.

Coding sample status was complicated by the common inclusion of cohabiting couples, and many samples comprised a mix of dating, married, and cohabiting couples. If the sample included a majority of married or cohabiting individuals, we coded it as married, and if it included a majority of individuals as dating and not living together, we coded it as dating. However, this coding should be regarded as indicative only. Relationship length for married couples provided another problem, as it is typically given as the length of the marriage, whereas relationship length for cohabiting or dating couples is always the actual (self-reported) length of the relationship. Thus, for the marital studies, we added 2.5 years to the numbers reported. This estimate was derived from several studies that have reported the length of time in the romantic relationship prior to marriage.

If the results from couples were reported for each gender separately, gender was treated as a subgroup within the study. For those studies that collected data from both men and women, but combined the results across gender, the sample sizes were doubled, before entering the data into the meta-analysis program, so that all studies were placed on an even playing field, given that effect sizes are weighted for sample size in the meta-analyses (see note in Table 2).

The nature of the sample (Item i above) was coded as clinical if the sample was selected according to the possession of high levels of depression, criminal convictions, high levels of relationship violence, and so forth. The community and college (university) labels are self-explanatory. None of the results varied significantly as a function of this factor, so it is not mentioned further.

The judgment category (Item k above) was coded into six subcategories: (a) judgments of partner personality traits; (b) judgments of negative attitudes, beliefs, or behaviors specifically directed to the partner or relationship (e.g., aggression, criticism); (c) judgments of positive attitudes, beliefs, or behaviors that are specifically directed to the partner or relationship (e.g., love, affection); (d) mind-reading judgments during interaction; (e) memories of past events or relationship states; and (f) predictions of future events or relationship states such as relationship satisfaction or longevity.

The entire set of categories was initially coded by the first author and tested for reliability by a second coder (the second author), excluding entering or calculation of the effect size (which was initially carried out and checked by the first author). The coders achieved 94% agreement across categories. Disagreements were resolved in discussion by the two coders. The first author entered the required effect size information into Comprehensive Meta-Analysis (CMA; see below) and checked for consistency. The details of all the studies included in the meta-analyses are described in Table 2.

Effect Sizes and Statistical Analyses

We used CMA (Version 2; Borenstein, Hedges, Higgins, & Rothstein, 2005) to analyze the data. Effect sizes were weighted for sample size with Hedges's *g*. CMA has the capacity to calculate effect sizes from appropriate raw data and to convert many kinds of data or effect sizes to the same metric. All effect sizes were converted to correlations with CMA, if not already in this metric.

Mean bias difference effect sizes were generally calculated within CMA from two sample mean scores—the judgment and the benchmark—along with other information (sample sizes, standard deviations, etc.). Positive mean-levels of bias were coded from the perspective of the judges. Thus, rating the partners as more attractive than the self-judgment of the partners, or recalling the relationships as improving over time more than they actually did, was coded as a positive bias. In the case of the positive interaction traits, if the judges rated the partners as, for example, loving them less than the partners reported, this was counted as a negative effect size. Similarly, if the judges rated the partners as criticizing them more than the partners reported, then this was coded as a

negative effect size. To calculate absolute levels of mean-level bias, we simply coded all effect sizes as positive.

Most of the tracking accuracy effect sizes (78 studies) were reported as correlations and simply entered into CMA in this format. Six studies (using nominal data) reported tracking accuracy results using chi-squares. For these studies, we either entered the effect sizes reported in the articles or calculated phi for those not reporting effect sizes (all these studies used 2×2 contingency tables). Where the results from specific studies were reported across several judgments (a common situation), the results were averaged across the judgments to produce one effect size for each study.

The most difficult set of studies to analyze were the 14 mind-reading studies. These studies always reported results in terms of percentage accuracy, and none reported effect sizes for overall accuracy (instead the focus of these studies was directed toward correlates or moderators of accuracy). These studies also varied in terms of whether they controlled for chance agreement. For studies, as the one described previously by Kahn (1970), it is easy to adjust the results for chance agreement because, as in Kahn, for example, the participants had a 1-in-3 chance of being accurate in judging the intention behind the emotional tenor of the message being sent by their partner.

Other studies used a procedure initially devised by Ickes et al. (1990) and modified by Thomas, Fletcher, and Lange (1997), in which each partner in a romantic relationship reviews videotaped relationship interactions (typically problem-solving discussions) and reports what he or she believes the other partner was feeling or thinking at specific points in the tape. Raters then later judge the accuracy of such reports by comparing these reports with their partners' self-reports of their own cognition that had been generated from the same videotapes. Some studies of this type adjust for chance agreement, but most do not. However, two studies have reported the chance accuracy rates by randomly combining self- and partner reports and calculating the overall percentage hit rates (Ickes et al., 1990; Thomas et al., 1997). On the basis of these two sets of results, we estimated the chance agreement as 4%, and we adjusted the rates accordingly for those studies reporting unadjusted rates.[7] The mind-reading effect sizes were analyzed by initially calculating one-sample *t* tests, comparing the adjusted rates to zero accuracy, and finally converting the effect sizes to correlations prior to inclusion in the CMA program.

The example given in the introduction (see Table 1) calculated tracking accuracy across a sample using a given trait. However, another method involves calculating such correlations within the couple (or participant) across a given set of items. The main advantage of using such a method is that it provides a convenient way of analyzing the associations between these idiographic measures of accuracy and other variables (such as relationship quality). Nevertheless, assessing the overall levels of tracking accuracy for each sample (using this latter method) still involves calculating effect sizes at the sample level, which is appropriate given that mean-level bias is also calculated at the sample level. There are advantages and disadvantages with each method, and it is possible for technical reasons (see Bernieri, Zuckerman, Koestner, & Rosenthal, 1994; Kenny & Winquist, 2001) to get different results at the sample level, depending on which method is adopted.

For the 98 studies analyzed in the current meta-analysis that assessed tracking accuracy, 75 used the method adopted in our example in the introduction (see Table 1), four reported results using both methods (which we combined and averaged for the meta-analyses), and five reported results using only the within-couple correlational method. The remaining 14 studies assessed mind reading; consistent with our prior description of these studies, this methodology is analogous to the within-couple method described previously. We combined the results in our meta-analyses across these two methods, but we also carried out exploratory analyses to test whether the results for tracking accuracy varied depending on the technique adopted to calculate tracking accuracy at the sample level.

In all cases in which the means being compared were dependent, we followed the advice of Dunlap, Cortina, Vaslow, and Burke (1996) and used the original standard deviations to calculate the pooled standard deviation (rather than the dependent *t* value). Random rather than fixed effects were calculated to test the overall effect sizes across studies, which is the more conservative approach and assumes that the data from different studies may come from different populations. CMA also generates 95% confidence intervals (CIs) around point estimates of effect size. For the analysis of moderating variables (such as gender, which formed subgroups within samples) or analyzing results across the six research categories (see above), we used a mixed-effects analysis. The *Q* statistic was used to test for the significance of group moderators (e.g., gender, judgment category).

Three estimates of publication bias were used. The funnel plot graphs the effect sizes against the inverse of the standard error, which should show higher variation with lower sample sizes but be symmetrical around the mean effect size (in the absence of publication bias). Duval and Tweedie's (2000) trim and fill analysis, which assumes there should be symmetry around the mean effect size, estimates how many studies are missing and what the point estimate would be if missing studies that reported null or low effect sizes were imputed. We used the random model for this analysis. Finally, the classic fail-safe test calculates the number of missing studies with effect sizes of zero needed to produce an overall, nonsignificant effect size (using fixed effects).

Results

Tracking Accuracy in Relationship Judgments

As expected, across the 98 studies for which we assessed tracking accuracy, the mean effect size (using a random-effects analysis) was significant and substantial ($r = .47$, $z = 23.85$, $p < .001$, 95% CI [.44, .50]).

We first tested whether these results for tracking accuracy were a function of the method used to measure accuracy. As noted previously, of these 98 studies, 75 assessed accuracy using correlations of specific traits across samples, whereas 23 used correlational or couple-centered techniques to produce scores that indexed accuracy within the couple or individual. The mean effect

[7] It might be argued that this particular method of assessing mind reading mixes up mean-level bias with tracking accuracy. However, the coding of participant-generated reports in this research paradigm is focused on the content rather than the affective tone; thus, we think it is reasonable to include this work under the umbrella of tracking accuracy.

size for studies that used the within-accuracy method ($r = .52$) was not significantly different ($p = .12$) from those that used the across-sample method ($r = .46$). In this analysis we used the within-accuracy results as reported by the four studies that reported both kinds of result. When we altered this analysis, so that we either used the across-sample effect sizes from these four particular studies or omitted them, the results were unchanged.

The results for tracking accuracy across the six judgment subcategories are shown in Table 3. The differences across the six subcategories were significant under a mixed-effects analysis, $Q(5) = 33.61$, $p < .001$. As can be seen, the effect sizes for tracking accuracy varied, with the highest effect size attained for memories of prior events or states ($r = .62$) and the lowest for predictions of future states or events ($r = .36$). Nevertheless, all effect sizes were positive, significant, and substantial.

In summary, there was strong evidence across 98 studies and six judgment categories that individuals produced moderately strong levels of tracking accuracy of their partners and relationships. Moreover, the same results were found regardless of the method used to assess tracking accuracy.

Mean-Level Bias in Relationship Judgments

As expected, across the 48 studies for which we assessed positive mean-level bias against some objective benchmark, the mean effect size (using a random-effects analysis) was significant and positive ($r = .09$, $z = 3.51$, $p < .001$, 95% CI [.04, .13]). Not surprisingly, when all effect sizes were converted to absolute mean-level bias effect sizes (simply indicating the amount of bias, either positive or negative), the mean effect size was higher and also significant ($r = .18$, $z = 9.54$, $p < .001$, 95% CI [.14, .21]).

The results for mean-level bias across the six judgment subcategories are also shown in Table 3. The effect sizes for mean-level bias were not significantly variable across judgment categories for the absolute index of mean-level bias (ranging from .12 to .27), when using a mixed-effects analysis ($p = .28$). However, the results were significantly variable across categories of positive mean-level bias, $Q(5) = 41.86$, $p < .001$. The reason for this difference in findings (see Table 3) is that, as predicted, both positive and negative interaction traits revealed a tendency to produce overall signed mean-level biases that were negative, and significant at the $p < .05$ level in the case of positive interaction traits.

In summary, a reliable tendency to produce both absolute and positive mean-level bias was found across 48 studies. However, one marked exception to this trend was studies investigating judgments of the interaction traits, either positive (e.g., love) or negative (e.g., criticism). Notably, as predicted, signed mean-level bias in both cases was negative. These results are consistent with error management theory (Haselton & Buss, 2000). Reassuringly, these particular results also mirror those from a prior meta-analysis of studies in the aggression domain reported by Archer (1999).

Publication Bias

To test for evidence of publication bias, we used three commonly used methods: funnel plot graphs, Duval & Tweedie's (2000) trim and fill analysis, and the classic fail-safe test. The funnel plot graphs of the tracking accuracy and positive mean-level bias studies are shown in Figure 2. As can be seen in both cases, the graphs show the standard result that precision tends to decrease with lower sample sizes. Publication bias would result if studies with smaller sample sizes (less precision) were more likely to be published if they attained positive and significant results. Duval and Tweedie's trim and fill analysis provides a quantitative estimate of this pattern by assuming that there should be symmetry around the mean effect size (if there is no publication bias) and then calculating what the overall effect size would be if missing studies that reported null or low effect sizes were imputed.

Under the random model, this analysis suggested that no studies on the left of the distribution were missing for either analysis. Finally, the fail-safe test indicated that 1,449 studies showing null results would need to be added to the positive mean-level bias studies, and 2,631 null studies would need to be added to the

Table 3
Effect Sizes and Results for Meta-Analyses of Tracking Accuracy and (Absolute and Positive) Mean-Level Bias Across Six Judgment Categories

Accuracy and bias	Number of studies	Effect size	CI (95%)	*z*	*p*
Tracking accuracy					
Personality traits	28	.43	[.38, .48]	14.35	<.001
Negative interaction traits	20	.51	[.46, .57]	15.33	<.001
Positive interaction traits	16	.40	[.33, .47]	10.01	<.001
Mind reading	14	.58	[.51, .64]	13.91	<.001
Memories	6	.62	[.52, .70]	9.77	<.001
Predictions	14	.36	[.28, .44]	8.28	<.001
Mean-level bias					
Personality traits	16	.15 (.18)	[.09, .22] ([.12, .24])	4.45 (5.70)	<.001 (<.001)
Negative interaction traits	10	−.07 (.13)	[−.15, .02] ([.05, .21])	1.51 (3.19)	.13, *ns* (<.002)
Positive interaction traits	7	−.11 (.15)	[−.21, −.002] ([.05, .25])	2.00 (2.91)	<.05 (<.005)
Mind reading	3	.09 (.12)	[−.07, .25] ([−.03, .26])	1.12 (1.61)	.26, *ns* (.11, *ns*)
Memories	4	.26 (.27)	[.12, .39] ([.13, .39])	3.57 (3.83)	<.001 (<.001)
Predictions	8	.23 (.25)	[.13, .33] ([.15, .35])	4.26 (4.92)	<.001 (<.001)

Note. For mean-level bias, the results for positive (signed) mean-level bias are presented first, and the results for absolute (unsigned) mean-level bias are in parentheses. CI = confidence interval.

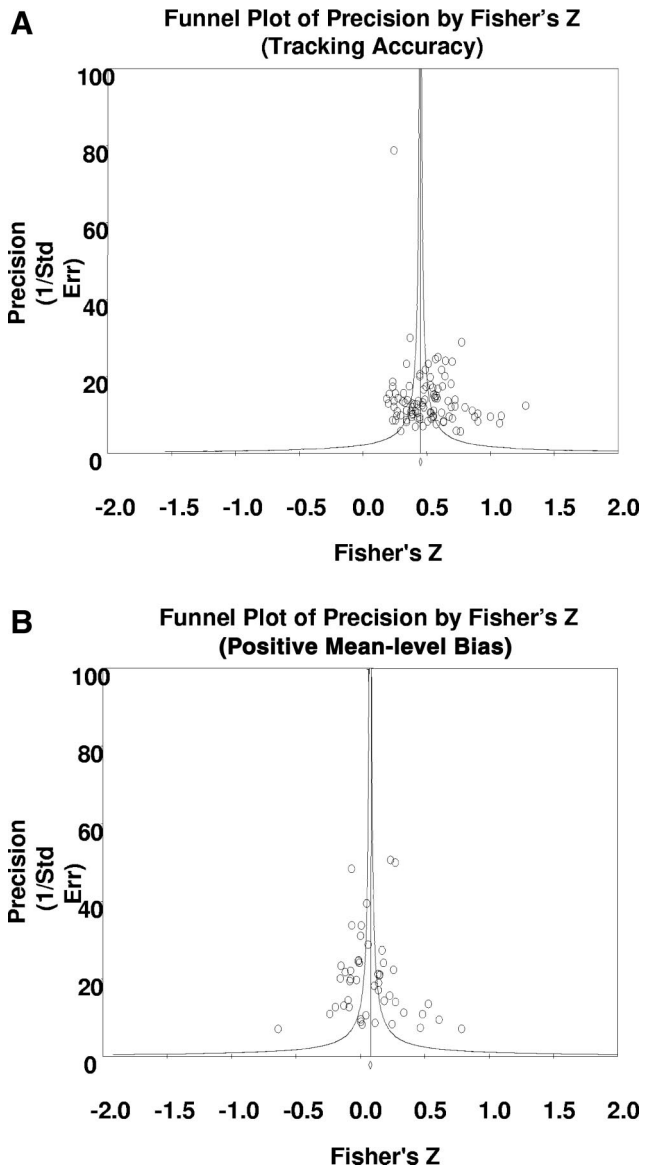

Figure 2. Funnel plots of precision for tracking accuracy (A) and positive mean-level bias (B). Std Err = Standard Error.

tracking accuracy set of studies, to produce nonsignificant results. We have not shown the distribution for the absolute mean-level bias findings, but the results revealed the same pattern, with no studies missing on the left of the distribution of the trim and fill analysis and 4,764 null studies required to produce a nonsignificant mean effect size. Taken together, these findings suggest that the results are reliable and not likely to be a product of publication bias.

The Independence of Mean-Level Bias and Tracking Accuracy

To handle this issue, we identified 38 studies from which we could obtain effect sizes for both positive mean-level bias and tracking accuracy. Thus, these studies had the same variables and samples involved in producing estimates of these two factors. Next, we calculated the mean effect sizes, using meta-analyses in the same way as previously, for both factors. As we had hoped, these effect sizes were similar to the full set of studies described above: $r = .07$ for positive mean-level bias and $r = .48$ for tracking accuracy. Finally, we correlated the two sets of effect sizes across samples, which produced a correlation of $-.005$ ($p = .98$, *ns*). When we recalculated this correlation by using the absolute levels of mean-level bias (rather than the signed levels as previously), the correlation remained weak and nonsignificant ($r = -.11$, $p = .53$).

Overall, these results suggest that the levels of mean-level bias and tracking accuracy attained by people are not merely conceptually and methodologically independent, as we have shown, but also empirically independent.

Links Between Positive Mean-Level Bias and Tracking Accuracy With Relationship Quality

We initially identified studies from which we could assess effect sizes for either the association between relationship quality and positive mean-level bias (14 studies) or the association between relationship quality and tracking accuracy (27 studies). The effect sizes for each set of studies were subjected to a meta-analysis, under the same approach as previously outlined with a random-effects analysis.

The results of the meta-analysis for the correlations between relationship quality and positive mean-level bias (14 studies) revealed that perceptions of higher relationship quality were significantly associated with higher amounts of positive mean-level bias ($r = .36$, $p < .001$). In contrast, for the meta-analysis of the correlations between relationship quality and tracking accuracy (27 studies), perceptions of relationship quality were unrelated to amounts of tracking accuracy ($r = .03$, $p = .42$). This difference in effect sizes was statistically significant, $Q(1) = 30.98$, $p < .001$. This result suggests that positive mean-level bias is hooked into the evaluative system of relationship cognition, whereas this does not seem to be the case for tracking accuracy.

The same assessments of publication bias were used as previously with these data. The funnel plot graphs showed little evidence of publication bias, with the effect sizes looking symmetrical around the mean effect size. Duval and Tweedie's (2000) trim and fill analysis confirmed this interpretation, showing that when three cases were imputed on the left side of the distribution for positive mean-level bias, the effect size remained positive and substantial ($r = .28$). Two studies were also imputed to the left of the distribution for the tracking accuracy analysis, but the resultant estimated effect size was even closer to zero than the original finding ($r = .006$). Finally, the fail-safe test (using a fixed-effects analysis) indicated that 923 studies showing null results would need to be added to produce an overall nonsignificant result for the set of mean-level bias studies, whereas only six studies would be required for the tracking accuracy set of studies to achieve a nonsignificant result. Taken together, again, these results suggest that the results are reliable and unlikely to be a product of publication bias.

The Moderating Variables of Gender and Relationship Length

Gender. We first tested the extent to which gender moderates overall levels of mean-level bias and tracking accuracy. Many of the studies examined included couples, so where possible we gathered effect sizes separately from studies according to gender. We also included male and female samples from studies that included either men or women.[8] Thus, the studies that summarized results across couples were omitted in these analyses. We located 34 usable studies for mean-level bias and 70 studies for tracking accuracy. Gender failed to exert a significant moderating influence on the overall results (using a mixed analysis), for either tracking accuracy (*r*s = .47) or positive mean-level bias (women: $r = .08$, men: $r = .07$). However, women did attain significantly higher levels of absolute bias than men (women: $r = .20$, men: $r = .12$; $p < .05$).

We next compared gender across studies according to two critically different kinds of judgments that were assessed in different studies: interaction trait attributions (both positive, such as partner love, and negative, such as partner criticism) and individual-level attributions that are more general in their focus, such as attractiveness or memories or predictions concerned with the self (assessed in the four categories of personality, mind reading, memory, and prediction). Our reasoning, laid out in the introduction, was that if women are more relationship oriented than men, and are therefore more sensitive to the cost and reward contingencies (as identified in error management theory), then this should lead women (more than men) to produce strong positive mean-level bias for the noninteraction categories (such as personality attributions). In contrast, the gender differences should reverse for the interaction traits because the relationship costs probably increase for exaggerating the positivity of such traits.

As can be seen in Figure 3, the predicted effects were obtained for the positive mean-level bias results: For studies measuring individual-level attributions, women produced more positive mean-level bias ($r = .21$) than men ($r = .10$), $Q(1) = 4.08$ $p = .04$. This result replicates Gagné and Lydon's (2003) results. However, as expected, for those studies that assessed interaction traits, the women produced a stronger overall negative mean-level bias ($r = -.13$) than men ($r = -.03$), although this difference was not significant ($p = .12$; sample sizes for these analyses ranged from 11 to 20). To test discriminant validity, we ran the same analysis for tracking accuracy. As expected, we found nonsignificant and slight differences between male and female samples regardless of whether individual-level or interaction-level traits were being rated (sample sizes for these analyses ranged from 23 to 42; see Figure 3).

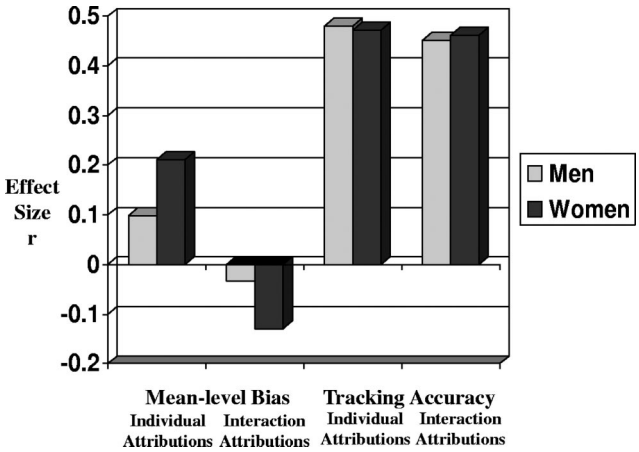

Figure 3. Gender differences in positive mean-level bias and tracking accuracy comparing effect sizes for individual-level attributions (Judgment Categories 1, 4, 5, and 6 in Table 2) with interaction attributions (Categories 2 and 3).

As noted previously, Gagné and Lydon (2003) also found preliminary evidence that women are more strongly oriented to relationships than men, and thus arguably do not need to make the shift to high levels of commitment to display partner-serving illusions to the same extent as men. We tested their explanation in the current study by calculating the link between relationship evaluations and positive mean-level bias as we did previously, but for men and women independently, using the subset of six studies that included such results for both genders. The results were consistent with their explanation, with the mean effect size being larger for the men ($r = .35$) than the women ($r = .21$), although this difference was not significant (perhaps not surprisingly, given the low power of this analysis).

Relationship length. We expected that the link between positive mean-level bias and perceptions of relationship quality should reduce in size as the length of the relationship increases across samples. To test this prediction, we found 13 studies from which we could derive both estimates. The mean length of time in relationships across these studies varied from 1.17 to 43.88 years ($SD = 12.32$). We used CMA to perform a metaregression analysis in which the sample-weighted effect sizes for mean-level bias (transformed to Fisher's *z*) were regressed onto the log-transformed measures of relationship length. In keeping with the recommendations of Borenstein, Hedges, Higgins, and Rothstein (2009), and our prior analyses, we used the random-effects model (method of moments). As can be seen in Figure 4A, this prediction was supported with the unstandardized slope being significantly negative ($b = -0.29$, $p < .02$, $R^2 = .23$). This finding is consistent with our reasoning that passionate love is more strongly linked to positive mean-level bias than companionate love.

We also predicted that the association between tracking accuracy and perceptions of relationship quality should go up as relationship length increases. To test this prediction, we used the same kind of analysis as previously, with the 27 studies from which we were able to collect both the relevant effect sizes and the measures of relationship length. The mean length of time in relationships across studies in this set varied from 1.00 to 43.88 years ($SD = 8.74$). However, as can be seen in Figure 4B, this

[8] When we repeated all the analyses examining gender as a moderating variable, but included only those studies that contained men and women subsamples, the results were very similar.

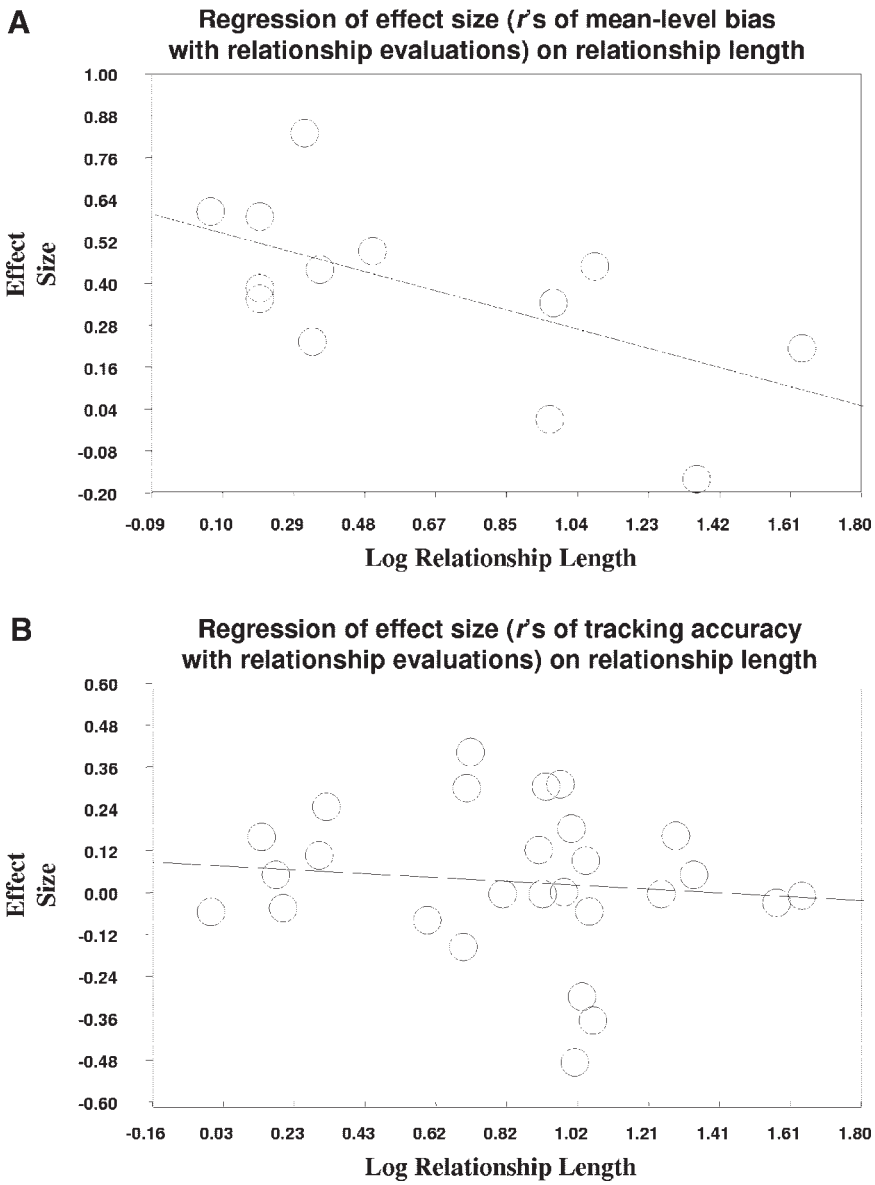

Figure 4. Scatterplots showing regression lines from the metaregression analyses in which the effect size variable representing the link between positive mean-level bias and relationship evaluations was regressed on the log of relationship length across samples (A), and the effect size variable representing the link between tracking accuracy and relationship evaluations was regressed on the log of relationship length across samples (B). Effect size correlations were converted to Fisher's *z*.

prediction was not confirmed, with a nonsignificant, almost flat slope produced ($b = -0.06$, $p = .42$).[9]

Associations Among the Moderating Variables and Relationship Satisfaction

Finally, we calculated associations among both moderators and relationship satisfaction. We located 22 studies that reported relationship satisfaction for men and women separately and found that the (standardized) mean scores were virtually identical and nonsignificant across studies according to a dependent *t* test ($p = .41$).

[9] Using the same approach, we assessed whether relationship length moderated the overall amounts of positive or absolute mean-level bias and tracking accuracy in a linear fashion. In all cases, the meta-regression analyses revealed relatively flat lines, and neither factor varied significantly as a function of relationship length. In addition, we entered all the relevant data into SPSS to test whether the slopes in any of these cases might be curvilinear, using multiple regression. We found no evidence in any of the analyses for curvilinearity.

We also located 34 studies that allowed us to calculate the link across studies between relationship satisfaction and (log-transformed) relationship length (both at the couple level). There was evidence that longer relationships were less satisfied, but the correlation was low and nonsignificant ($r = -.19$, $p = .29$). Overall, these results suggest that these associations would be unlikely to explain any of the findings reported previously.

Discussion

In the Discussion we first summarize the results, then discuss the role of the moderators listed in Figure 1. Next, we discuss the role of projection and the research using more complex path-analytic models. We then consider what the research tells us about judgments in the early stages of mate selection, and the issue of what people know and want in intimate relationships in terms of mean-level bias and tracking accuracy. We attempt to draw conclusions in each case in terms of the model shown in Figure 1. We finally discuss some implications for the wider study of motivated cognition and rationality in social cognition, as well as some new directions for research, and draw some final conclusions.

Positive Mean-Level Bias and Tracking Accuracy in Romantic Relationships

The overall results showed that across 98 studies and 27,064 individuals, the tracking accuracy obtained was reliable and substantial ($r = .47$). The overall amounts of positive mean-level bias were lower, but were also reliable across 48 studies and 9,393 individuals ($r = .09$). These findings are generally consistent with those of other, considerably more focused and limited qualitative literature reviews, but they provide the first evidence from a systematic meta-analysis of work across many domains in the field, and various novel findings were also produced. Moreover, there was no evidence that the results were a function of the method adopted to calculate tracking accuracy (within-couple indices or traits across samples). And the research surveyed used a tremendous range of judgments (see Table 2), which suggests that the results are generalizable across many judgmental domains.

The overall effect findings varied significantly across the judgment categories. For the tracking accuracy findings, the best performing category comprised memories of past events or relationship states ($r = .62$), and the lowest was predictions of future relationship events or states ($r = .36$). Perhaps not surprisingly, people are more accurate at recalling the past in romantic relationships than predicting the future. Nevertheless, as expected, the levels of tracking accuracy were robust and significant for all judgment categories (see Table 3). The pattern of findings across categories was also quite variable for positive mean-level bias across categories (see Table 3 and Figure 3), with moderate levels of positive bias produced for personality traits, memories, and predictions, but with negative mean-level bias obtained for judgments of the partners' beliefs, attitudes, or behaviors directed to the partner or relationship (the interaction traits).

This latter pattern of findings for mean-level bias is consistent with error management theory's claim that this kind of bias is likely to be a function of the (expected) costs and rewards involved (Haselton & Buss, 2000). Research has shown that individuals in romantic relationships are exquisitely attuned to behavior emanating from their partners that communicates attitudes and beliefs about them (so-called reflected appraisals; Murray, Bellavia, Rose, & Griffin, 2003; Overall & Fletcher, in press). Thus, when judging the extent to which their partner forgives them, loves them, or is angry and aggressive toward them, people may be on a relative hair trigger, compared with the other kinds of judgments. The reason for this may be that the failure to notice and react appropriately to such partner behavior is likely to have deleterious effects on the relationship.

Take forgiveness (an example of a positive reflected appraisal). If Mary optimistically (and wrongly) assumes that her partner has forgiven her infraction, this may lead her partner to desire payback (see Fitness & Peterson, 2008) and to perceive Mary as insensitive or self-absorbed. It may also fail to motivate Mary toward reconciliation and relationship repair. In contrast, perceiving one's partner as more attractive, kind, and ambitious than he or she actually is seems, at face value, to possess few of these downsides.

On the basis of our analysis in the introduction, mean-level bias and tracking accuracy of judgments in intimate relationship contexts are conceptually and methodologically independent. The novel results from the meta-analyses showed clear and consistent evidence that these two forms of accuracy are also empirically independent. First, across the 38 studies that reported findings for both mean-level bias and tracking accuracy, the two effect sizes were unrelated. Second, more positive mean-level bias in relationship judgments was significantly associated with more positive perceptions of relationship quality across studies ($r = .36$), whereas the amount of tracking accuracy generated was unrelated to perceptions of relationship quality ($r = .03$). Third, gender moderated positive mean-level bias in certain judgment categories in predictable ways but failed to moderate tracking accuracy. Finally, the linear association between positive mean-level bias and perceptions of relationship quality significantly diminished across samples as a function of increasing relationship length, whereas the corresponding linear relationship for tracking accuracy was relatively flat and not significant.

As an illustration of some of these findings, consider some recent research on the so-called affective forecasting error in relationship contexts (Eastwick, Finkel, Krishnamurti, & Loewenstein, 2008). Prior evidence has indicated a robust tendency in nonrelationship contexts for people to predict greater levels of negative or positive affect, following negative or positive events, than actually eventuate (an example of mean-level bias). The research by Eastwick et al. (2008) found the same effect when individuals first predicted, then experienced, the affective outcomes associated with a dating relationship breakup; people experienced significantly less distress than they predicted concerning the relationship breakup (effect size $r = .66$). However, they also evinced significant tracking accuracy of their emotional reactions ($r = .44$). And the forecasting (mean-level) error disappeared for those who were not in love with their partners, indicated that it was likely they would start a new romantic relationship a few weeks prior to the breakup, or initiated the breakup. In short, only those who were significantly invested in the relationship predicted more distress than they experienced when the relationship actually dissolved. It is hard to resist the interpretation that this particular bias has a functional basis, given that it should motivate individuals who have much at stake to maintain and improve their romantic relationship, and perhaps retain their mates.

The overall results generally support our prior suggestion that the levels of positive (or negative) mean-level bias in relationship judgments are tied into pivotal elements of local relationship theories that are evaluative and affect laden, the prototypical example being relationship evaluations. Of course, the results from the meta-analysis do not speak to what causes what. However, Murray and her colleagues (Murray et al., 2002, 1996a, 1996b) have established that more positive bias in partner judgments predicts higher levels of relationship happiness over time (Murray et al., 2002; Murray et al., 1996a) and lower rates of dissolution (Murray et al., 1996b), and serves to mold the relevant self-perceptions of the partner in a positive direction (Murray et al., 1996b).

Other research suggests that more positive bias does not always produce beneficial consequences for the relationship. Steadfastly maintaining a sunny set of partner and relationship judgments works well when the problems are mild and both participants are socially skilled and committed. However, when relationships are in serious trouble, people seem to do better over time when, in this particular case, a more negative and realistic stance is adopted (McNulty & Karney, 2004; McNulty, O'Mara, & Karney, 2008).

A strong version of the love-is-blind thesis would posit that individuals in love should largely ignore incoming data (e.g., behavior from their partners) and rely more or less completely on their preexistent beliefs or lay theories (including relationship evaluations) to interpret negative events and crush doubt. Instead, the evidence suggests (consistent with our model in Figure 1) that relationship-quality judgments influence positive mean-level bias in relationship judgments and vice versa.

Moderators of Mean-Level Bias and Tracking Accuracy in Romantic Relationships

This meta-analysis explicitly addressed two of the moderators listed in Figure 1, relationship stage and gender, which we next discuss. However, we also discuss the other moderators listed in Figure 1 and mention related research.

Relationship stage. Simply being in a happy, romantic relationship seems to automatically generate a positive bias. However, as our model postulates, various moderating factors can prime and increase the power of different goals including the need to maintain relationship satisfaction. The stage of the relationship is one such important moderating variable. Overall amounts of positive mean-level bias did not significantly decline according to the length of the relationship.[10] However, as noted, the link between positive mean-level bias and relationship satisfaction markedly decreased as the mean length of the relationship increased across studies (see Figure 4).

Our interpretation of this finding is that passion and romantic love strongly push judgments into the rose-tinted realm. This interpretation is consistent with an evolutionary account of romantic love as a device to encourage long-term bonding in mates (Gonzaga & Haselton, 2008), and with the developing neuropsychological evidence concerning how cognitions and emotions associated with romantic love are associated with specific neuropeptides, neural networks, and regions of the brain (Fisher, Aron, & Brown, 2006). On this account, as romantic love and obsession cool and change to a companionate form of love marked by contentment and commitment, the motivation that produces positive mean-level bias also weakens.

Like mean-level bias, the levels of overall tracking accuracy across studies did not change as a function of relationship length across studies.[11] This finding indicates that accuracy probably peaks relatively early in romantic relationships. In contrast to the case for positive mean-level bias, however, the effect sizes representing the link between tracking accuracy and relationship satisfaction did not change as a function of relationship length across studies (see Figure 4). Thus, in spite of our prediction, there was no tendency for tracking accuracy to enhance relationship satisfaction over time in romantic relationships.

Nevertheless, it seems likely that certain stages of the relationship might prime the need for more accurate predictions of the future of the relationship. Fletcher and Thomas (1996) originally proposed that the goals of producing realistic (minimal positive bias) and accurate (good tracking accuracy) predictions and evaluations concerning the relationship might be especially salient when important decisions regarding changes in commitment are being made (e.g., when people are deciding whether to leave the relationship, move in together, and get married). In contrast, once important decisions have been made concerning relationship investment, the goal of maintaining relationship satisfaction should dominate, leading to positively biased processing once again taking center stage.

Testing these ideas, Gagné and Lydon (2003) put their participants into either deliberative (postdecisional) or implemental (predecisional) mind-sets. Work by Gollwitzer and colleagues (for a review, see Gollwitzer, 2003) had found that adopting a deliberative mind-set (compared with an implemental mind-set) is associated with more even-handed and impartial processing of information and less optimistic perceptions and predictions about the self. Across three studies (two correlational and one experimental), Gagné and Lydon found that individuals who were encouraged to think (or were already thinking) in an even-handed, predecisional fashion were quite accurate in predicting the long-term demise of their relationship ($r = .67$), whereas those who were pushed away from the goals of prediction and truth-seeking by virtue of being in a postdecisional mental set were not very accurate in their predictions ($r = .19$). Thus, in terms of our model in Figure 1, the demands of the social context seem to have the capacity to push people toward or away from goals such as prediction or truth that then may influence tracking accuracy.

Gender. Women produced slightly more tracking accuracy than men, but this difference was small and nonsignificant, regardless of the judgment domain. There is evidence that women generally have more social or relational intelligence than men (see Fletcher, 2002) and that women read nonverbal behavior more accurately than men (Hall, 1978), so this null finding is puzzling. It has been argued that gender differences (e.g., in mind reading) are mainly a function of motivation rather than ability, and there is evidence that the superiority of women over men in mind-reading

[10] See Footnote 9.

[11] See Footnote 9.

strangers disappears when motivation is enhanced (Klein & Hodges, 2001). Still, higher motivation levels for long periods of development over the life span are likely to result in permanent increases in any ability, whether it is playing chess, drawing pictures, or reading minds.

Our conclusion, based on these results, is that both men and women share similarly high levels of motivation to understand and predict important features of potential or actual long-term romantic partners over time and, thus, do not differ in their success at achieving tracking accuracy, regardless of the kind of judgment being assessed.

We found, however, that women overall produced significantly higher levels of absolute bias than men. Moreover, confirming research by Gagné and Lydon (2003), we found that women demonstrated significantly more positive mean-level bias than men (see Figure 3) when judging their partner on partner-level traits or mind readings that were not interaction traits (e.g., general personality traits). In contrast, women produced more negative amounts of mean-level bias than men (see Figure 3) for the interaction traits (a predicted reversal of the findings just described).

This latter finding should be treated with caution given that this gender difference was not significant. Nevertheless, the results overall are consistent with the proposition that women are more concerned than men with managing and monitoring intimate relationships, and are thus more keyed into the rewards and costs of adopting default mean-level biases. In short, the goal of maintaining relationship satisfaction is more salient and accessible for women than men (see Figure 1). In addition, this pattern of gender differences was lent some discriminant validity, given that there was little difference between male and female tracking accuracy when comparing individual-level with interaction attributions (see Figure 3).

Overall, then, in terms of the model (see Figure 1), these results suggest that gender interacts with the goal of attaining relationship satisfaction to produce gender differences in the amounts of both absolute and signed mean-level bias.

Judgment relevance. Consistent with our results and interpretations, some research suggests that the relevance of the judgments to romantic relationships (see Figure 1) moderates the links between positive mean-level bias or tracking accuracy and relationship-quality judgments. These findings show that mean-level bias is more positively biased, but tracking accuracy is higher, when judgments are on dimensions such as warmth, attractiveness, and status. In contrast, both positive mean-level bias and tracking accuracy are diminished when traits that are less central to mate selection or the success of romantic relationships are being judged, such as extroversion and artistic ability (Boyes & Fletcher, 2007; Gill & Swann, 2004). Thus, in term of our model (Figure 1), these findings imply that when the significance and relevance of the judgments to intimate relationships are high, this tends to engage goals such as evaluation and relationship satisfaction, which, in turn, influence the judgments advanced.

Individual differences. Individual differences as moderators in Figure 1 include dispositions that are not directly concerned with intimate, sexual relationships, including personality factors and self-esteem. For example, Murray, Holmes, and Griffin (2000) have shown that lower self-esteem is associated with more negative mean-level bias in terms of underplaying the amounts of love and satisfaction actually reported by the partner.

Recent diary studies by Murray and others (Murray, Bellavia, et al., 2003; Murray, Griffin, et al., 2003; Murray, Griffin, Rose, & Bellavia, 2006; Murray et al., 2000) also document the subtle and dynamic nature of these processes over short periods (typically 3 weeks) in romantic relationships. These studies suggest that when the partner is perceived as being insensitive or transgressing in some way, low self-esteem motivates withdrawal from the relationship, the production of uncharitable attributions, and a slide in relationship satisfaction.

In terms of the model in Figure 1, these results suggest that negative relationship behavior (a moderator) interacts with self-esteem (an individual difference moderator) to differentially prime the goals of evaluation and avoiding rejection,[12] which produce a more negative local relationship theory, and more negative mean-level bias associated with specific judgments. Crucially, the same studies show evidence of self-fulfilling prophecy effects so that the partners of such folk tend also to become disillusioned over time. We say *crucially* because these particular results could be obtained only if partners were (to some extent) accurately tracking behavior reflecting the dissatisfaction of their partners.

These results reflect the exquisitely fine-grained way in which partners depend on and influence each other in romantic relationships. They also illustrate the point that relationship interactions are shaped by, and have behavioral consequences as a function of, the mean-level bias and tracking accuracy attendant in relationship and partner judgments.

Social context. A growing body of research has implicated the importance that stressful local conditions play as a moderating variable in influencing bias and accuracy in relationship judgments. For example, Collins and Feeney (2004) used an experimental paradigm that manipulated messages of support by romantic partners prior to participants performing a stressful task (preparing and giving a speech that would purportedly be videotaped and rated by other students). Even when controlling for the actual quality of the support given, more anxiously attached adults were more biased toward perceiving their partners as less helpful and well intended.

J. A. Simpson and colleagues (J. A. Simpson, Ickes, & Blackstone, 1995; J. A. Simpson et al., 1999) designed a methodology in which individuals mind-read their partners in a threatening context by observing their partners rate the desirability of highly attractive opposite-sex individuals from a local dating pool. Overall, they found that those who were in closer and more intimate relationships were motivated to produce more *inaccurate* judgments of what their partners were thinking and feeling. However, more anxiously attached individuals were also more accurate in their mind readings, were correspondingly more distressed by the experience, and tended to suffer a greater loss of confidence in their relationships.

In terms of the model in Figure 1, these studies suggest that stress (a moderating variable) interacts with attachment working models (general relationship theory) to differentially prime goals such as evaluation and the need to protect the self. Specifically for individuals with anxious working models, this increases monitor-

[12] See Footnote 4.

ing, engages the resources of local stored relationship theories, and thus increases both negative mean-level bias and the levels of tracking accuracy.

The Role of Projection (aka Assumed Similarity) and Path-Analytic Models

As previously noted, projection of the self could be either an artifact or a heuristic that individuals use to attain accuracy. For example, it is possible that higher levels of relationship satisfaction are associated with either tracking accuracy or positive mean-level bias because happier couples tend to project more and also happen to be more similar. Alternatively, if couples are similar to varying degrees on certain traits or dimensions, then projecting appropriately on these traits should lead to more accuracy. What does the research say?

Consistent with our prior results, there is considerable evidence that individuals who perceive more similarity between themselves and their partners are indeed happier with their relationships (see Montoya, Horton, & Kirchner, 2008). However, as we know from the assortative mating literature, couples are quite often similar in many domains. Schul and Vinokur (2000) found that actual similarity across a wide range of domains (e.g., health, job stress, family life) was high (ranging from .40 to .72) and that the individuals' tendency to project in assessing their partners on such dimensions was correlated .33 with actual similarity. Thus, it is quite plausible that individuals may attain accuracy in their judgments by the judicious use of similarity judgments (projection).

Kenny and Acitelli (2001) outlined a commonly used model in which self- and partner ratings on specific dimensions comprise both the independent variables and the dependent variables. Figure 5 presents a generic example of such a model (termed the actor–partner independence model; see Kenny, Kashy, & Cook, 2006). This kind of model, often presented in elaborated versions with additional mediating variables, is commonly used in the relationship literature, typically in combination with statistical techniques such as structural equation modeling and hierarchical linear modeling. Paths a and b represent the extent to which both men and women project on a specific trait or dimension (controlling for actual similarity), and Paths c and d represent the accuracy paths (controlling for actual similarity).

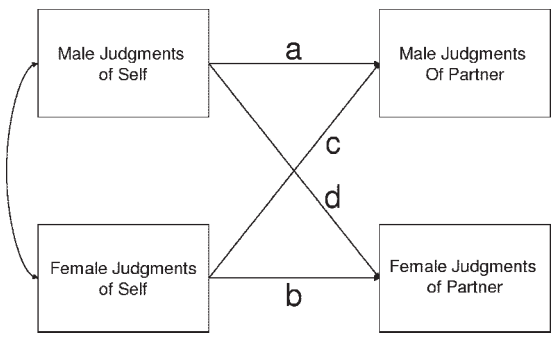

Figure 5. A generic example of the actor–partner independence model.

Note that all the effects analyzed in the meta-analyses (deliberately) assessed and used direct effect sizes that did not statistically control for other variables. Research that has calculated such accuracy paths controlling for actual similarity has produced mixed results, sometimes reporting quite large drops in the direct accuracy effects (e.g., Kenny & Acitelli; Kirchler, 1999; Sillars et al., 1984) and sometimes reporting that actual similarity (and therefore projection) played a relatively minor role (Saffrey, Bartholomew, Scharfe, Henderson, & Koopman, 2003; Wilhelm & Perrez, 2004).[13]

When accuracy and projection effects are calculated (while controlling for actual similarity), both effects are typically significant, with projection effects sometimes proving to be much bigger than accuracy effects (see, e.g., Murray et al., 1996b). But this may not be fair to the layperson. One obvious problem is that if people use their reasonably accurate perceptions of similarity to drive their partner judgments, then the accuracy paths will be underestimated. Indeed, Kenny and Acitelli (2001) showed that the indirect path to accuracy (through projection) can contribute substantial amounts of variance. For example, when their participants provided judgments of something intensely dyadic, such as the extent to which their partners enjoyed sex, considerable accuracy was attained but almost entirely as a function of individuals accurately (and appropriately) projecting their own levels of enjoyment onto their partners.

In interpreting these results in terms of the model in Figure 1, the self-judgments that are being projected could either be counted as a moderating variable (under the label of individual differences) or come under the banner of stored local relationship theories. We suspect that many of the self-judgments being projected in intimate relationships are essentially relational judgments, or are part of what Chen, Boucher, and Tapias (2006) called the relational self, which is the self in relation to significant others. Prototypical examples are interaction attributions of the sort we have already examined (e.g., relationship satisfaction) or factors that tend to be shared across partners (e.g., financial or housing problems).

An examination of the results produced by Kenny and Acitelli (2001) and Schul and Vinokur (2000) appears to confirm this suggestion, with projection reaching substantial proportions for attributions such as closeness, enjoying sex, family life events, relationship satisfaction, and financial strain but (sensibly) being relatively weak for health, job satisfaction, or hours spent at work. Thus, the extent to which tracking accuracy is obtained in specific judgments via projection is yet another example of the causal links

[13] In an analogous procedure, Thomas and his colleagues (Thomas & Fletcher, 2003; Thomas et al., 1997) developed a way of measuring projection in mind reading by getting partners to report what they were thinking and feeling at each point of the stopped video of the interaction. Raters then were able to assess actual similarity, projection (the authors used the term *assumed similarity*), and accuracy for each set of reports across each dyad. Projection (assumed similarity) scores did not correlate highly with mind-reading accuracy scores in the two studies using this procedure (Thomas & Fletcher, 2003; Thomas et al., 1997), and statistically controlling for projection did not change the results.

shown in Figure 1 between local relationship theories and specific judgments.[14]

What About the Early Stages of Mate Selection?

Little research has examined the accuracy or bias of individuals in the early stages of intimate romantic relationships, although some work has compared the accuracy of strangers' judgments with those of people involved in intimate relationships. However, there has been a considerable amount of research studying the tracking accuracy of judgments among (relative) strangers in nonromantic contexts, especially for the case of personality traits. This research suggests that straightforward personality judgments such as extroversion, conscientiousness, and agreeableness can be assessed with surprising accuracy (using self-other agreement) by strangers after receiving minimal information (for a recent brief review, see Beer and Watson, 2008). However, the best replicated finding across studies concerns the moderate ability of people to judge extroversion under conditions of minimal acquaintance (see Beer and Watson, 2008), which is often explained according to the greater observability of this trait compared with other personality traits (Funder & Dobroth, 1987).

From an evolutionary angle (as previously noted), it is important to know how accurate people are in assessing attributes that are pivotal in mate selection contexts, including physical attractiveness and status. There is scant research on the judgment of status in minimal-information studies, but one study showed that simply viewing photographs of university employees having a conversation with a coworker yielded remarkably accurate judgments of their relative work status ($r = .55$; Mast & Hall, 2004).

There is also extensive research on rating physical attractiveness in others, with good evidence that such judgments are quite accurate based on minimal observations or interactions (although this research is rarely carried out in mate selection contexts). For example, Marcus and Miller (2003) had participants rate their own physical attractiveness and that of other men and women who were sitting together in small groups, using a round-robin design. There was good consensus on the level of attractiveness for specific targets, and targets' self-perceptions generally matched well with how they were perceived (correlations ranging from .28 to .53). Moreover, individuals' metaperceptions of how they were perceived generally by others were accurate (with correlations ranging from .26 to .49). As the authors concluded, "Generally, we know who is handsome or pretty, and those who are attractive know it as well" (p. 334).

One important subset of studies has compared the accuracy of relationship outsiders (e.g., friends) with that of relationship insiders (i.e., the relationship partners themselves), with mixed results. MacDonald and Ross (1999), for example, found that college students made more positive but less accurate predictions about the longevity of their own current dating relationships than either their roommates or their dating partners' parents. However, this result was not replicated by Loving (2006), who also found (as did MacDonald and Ross) that the best predictors were the daters' self-reported levels of commitment.

To gain a more objective analysis of the results of these (published) studies, we carried out a meta-analysis of the seven studies that we found that had included insider and outsider judgments of the same relationships. The mean tracking accuracy of insiders (partners) was .58, for friends or family members it was .51, and for strangers it was .47. Although these differences were not significant across the observer types, the results are consistent with the research discussed above revealing that relationship insiders have the edge but that strangers can reach surprisingly high levels of accuracy.[15]

In summary, these findings and analyses in terms of our model (see Figure 1) imply that general relationship theories play a large role in judgments that occur early in mate selection contexts, given that local relationship theories are bound to be simple and underdeveloped. However, the dynamics in such contexts are complex, with evidence that women are more coy than men about their romantic interest and are thus harder to read by observers (Place, Todd, Penke, & Asendorpf, 2009), and women are more choosy than men and thus are inclined to demand more according to perceptions of their own mate value (Todd, Penke, Fasolo, & Lenton, 2007).

There is also some evidence that the press for consistency among general lay theories and perceptions is not especially strong in the very early stages of intimate relationships (Eastwick & Finkel, 2008; Fletcher et al., 2000). However, we think that individuals rapidly develop local relationship theories in the early stages of mate selection (see Fletcher et al., 2000) and that the goals of prediction and evaluation should be highly accessible and influential at this stage.

What Do People Want, and Do They Know When They Have It?

People express strong desires for authentic, open, and honest relationships (Fletcher et al., 1999). Research testing self-verification theory has consistently suggested that, at least for married couples, individuals in romantic relationships prefer to have their partners think of them in ways consistent with their self-perceptions, even when such attributions are negative (De La

[14] Yet another kind of "bias" initially posited by Cronbach (1955) is termed stereotype accuracy. This is a misleading term, as it is defined statistically in terms of the way in which raters in general judge different traits in different ways (rather than being equivalent to stereotypical beliefs per se). This factor can be assessed and controlled for in different ways, but the same argument has arisen regarding its use as with projection. That is, controlling for it removes the variance associated with a cause of accuracy or even a component of accuracy itself (see Funder, 1999). However, as pointed out by Kenny et al. (2006), examining such correlations with and without controlling for this factor can be informative. Only three tracking accuracy studies of the 98 included in our meta-analysis explicitly analyzed stereotype accuracy (Acitelli et al., 2001; Corsini, 1956; Swann & Gill, 1997), but all showed that stereotypes seemed to play some part in the production of overall tracking accuracy. Note, however, that this kind of bias applies only to within-couple profile correlations and does not affect across-sample trait measures (Kenny & Winquist, 2001). Thus, given that our tracking accuracy findings were not influenced by which correlational measure was used, this factor does not appear to have had a major inflating influence on our results.

[15] Of these seven studies, two assessed mind reading (Koerner & Fitzpatrick, 2002; Thomas & Fletcher, 2003), three assessed predictions of relationship breakups (Agnew, Loving, & Drigotas, 2001; Loving, 2006; MacDonald & Ross, 1999), and two used behavioral interactions (Collins & Feeney; 2000; Lorenz, Melby, Conger, & Xu, 2007).

Ronde & Swann, 1998; Swann, De La Ronde, & Hixon, 1994). However, there is also evidence that partners welcome positive mean-level bias, especially in judgments that are central to intimate relationships, such as warmth, attractiveness, and status (Boyes & Fletcher, 2007; Swann, Bosson, & Pelham, 2002).

This apparent contradiction is nicely resolved by the distinction drawn in the current article, and the associated proposal that intimates may prefer to have reflected appraisals that are both positively biased and high in tracking accuracy. This hypothesis has been tested by Lackenbauer et al. (in press), who experimentally manipulated the views that partners in romantic relationships had of each other in terms of two independent dimensions: high versus low mean-level positive bias and high versus low tracking accuracy. The results showed that the effects were additive, with the maximum satisfaction attained when the feedback from the partner was both positively biased and accurate.

A separate question concerns the extent to which individuals are normally aware of the amounts of mean-level bias and tracking accuracy in intimate relationships. Swann and Gill (1997) found that people were more confident of the accuracy of their judgments (using questions assessing a wide range of attitudes and behaviors) if their relationships were longer and the lay theories of their partners were more complex and integrated. Unfortunately, there was a minimal link between confidence ratings and accuracy. In a similar vein, Thomas et al. (1997) reported that there was no significant association between the accuracy of mind reading attained during problem-solving interactions and the confidence expressed in the attributions.

In contrast, there is evidence that people can accurately gauge the amounts of mean-level bias in their relationships. Barelds-Dijkstra and Barelds (2008) reported the familiar findings of overall positive mean-level bias when comparing attractiveness judgments made of relationship partners with the self-judgments of the same partners. However, when participants were asked to judge how attractive other people found their partner, the mean-level bias significantly dropped, suggesting that individuals were aware to some extent of the bias inherent in their own perceptions. More directly, Boyes and Fletcher (2007) found that men and women in romantic relationships were quite accurate in assessing the amounts of positive mean-level bias in warmth, attractiveness, and status, in terms of both their own judgments of their partners and the bias inherent in their partners' judgments of them.[16] Moreover, these effects were not a function of relationship satisfaction or self-esteem, and so were unlikely to be driven simply by halo effects.

Rusbult, Van Lange, Wildschut, Yovetich, and Verette (2000) showed that when people were given accuracy instructions, this weakened the amounts of positive bias (measured as a within-participant variable) when describing their relationships. This result suggests that people are able to control the production of mean-level bias, depending on which goals are primed. Moreover, the correlation between relationship commitment and positive bias was strongest in a threat condition (.61), moderate in the control condition (.37), and weakest when an accuracy goal was primed (.17). These results support our argument that more positive relationship evaluations motivate greater positive mean-level bias, but that the extent to which they do will be linked to the goals that are primed (see Figure 1).

Given the sparse research on these topics, caution should be exercised in drawing firm conclusions. However, the results do raise the question of why individuals may possess better metacognition about mean-level bias than tracking accuracy. Boyes and Fletcher (2007) argued there are two conditions for meta-accuracy of such judgments to flourish. First, there needs to exist a familiar, accessible, and relevant lay theory, such as (in this case) the notion that love is blind. Second, the existence of such biased judgments needs to be expressed and observable during behavioral interactions. Indeed, there is good evidence that partners in intimate relationships often try to change each other on central relationship dimensions such as warmth, attractiveness, and status (Overall et al., 2006, 2009), and couples will often work collaboratively to improve themselves, such as going on a diet or joining a gym together to lose weight. And it seems likely that individuals in romantic relationships may often communicate their respective (positively or negatively biased) views to one another verbally and nonverbally. If Mary asks Stephen if she looks fat in her new dress, Stephen's reply will be diagnostic about his (biased) perceptions of Mary's attractiveness. Thus, individuals in relationships are likely to have abundant evidence available to help judge the amounts of mean-level bias inherent in partner judgments of specific, relationship-central traits.

In contrast, developing accurate metaperceptions of the tracking accuracy of mind reading or personality traits may be more intractable, and perhaps beyond the cognitive resources of many individuals, because of the relative difficulty of the task. For example, with respect to mind reading, the quality of feedback may be poor (partly because individuals often mask their own emotions and cognitions); the processes involved are often rapid, fleeting, and unconscious; and finally, unlike with mean-level bias, the perceiver is not making relatively simple, relative assessments between a judgment and a benchmark of some kind (see Ickes, 1993).[17]

Regardless of such conjectures, important questions remain concerning the moderating conditions that might influence what people want, what they know about mean-level bias and tracking accuracy, and what the consequences for intimate relationships might be for both kinds of factors.

Rationality, Bias, and Motivated Cognition in Intimate Relationships

Our findings and analyses are consistent with the emerging consensus that cognition and motivation are interwoven (Dunning, Kunda, & Murray, 1999; Hassin, 2008). We agree with the claim often advanced that one of the prime motivating goals in intimate

[16] In contrast to the ability of individuals to accurately read the levels of positive bias in their partners' views of self, the apparent ability of individuals to accurately assess their own levels of positive bias in perceiving the partner as the target raises some thorny theoretical issues. Boyes and Fletcher (2007) argued that under certain circumstances, people are able to switch perspectives from cheerleader to truth seeker when making such judgments (see also Gagné & Lydon, 2004).

[17] This does not imply that tracking accuracy does not have consequences. People do not have to be consciously aware of processes, or make accurate meta-accuracy judgments, for tracking accuracy to exert causal influences.

relationships for many people is to build and maintain high levels of commitment and judgments of relationship quality, which, in turn, bias relevant judgments in a positive direction.

Results from the meta-analyses provide good support for this proposition. First, the results revealed a robust link between higher relationship satisfaction and more positive mean-level bias (using objective benchmarks). Second, as relationships became longer, and arguably slipped from a passionate to a more relaxed, less intense form of love, the link between relationship satisfaction and positive mean-level bias diminished. Third, the default mean levels of bias across samples tended to be positive, with the one exception of the interaction traits. However, note that we explained this apparent anomaly in terms of the dangers to the relationship of assuming that the partner is more positive about the self than is actually the case. Thus, if our explanation is correct, even this case can be explained in terms of the desire to protect and enhance the relationship (a motive that appears to be stronger for women than men).

However, this thesis does not imply that such lay relationship thinking is irrational. It is important to note that the way in which prior theories or beliefs should be (and are) adjusted in the light of new evidence is at the heart of the empirical study of rationality and bias in human judgment as well as the philosophy of science (Fletcher, 1995). Krueger and Funder (2004), along with many others, have argued against the use of null hypothesis testing and for its replacement with Bayes' theorem. Bayes' theorem encapsulates the key insight that not all theories are created equal. Accordingly, the power of data to disconfirm or confirm a particular theory is constrained by the a priori weighting assigned to a specific theory as likely to be true.

A major limitation of Bayes' theorem is that it not clear how the probability or weight should be assigned to the prior belief or theory (Haig, 2009). Nevertheless, a Bayesian approach makes it clear that using prior beliefs and theories to weight and evaluate incoming data is quite rational and appropriate. Or to put it another way, from a scientific perspective, prior lay theories (beliefs, expectations, etc.) should bias the way in which relevant data are interpreted and explained for the layperson and scientist alike.

Now, from a scientific angle, prior theories are always considered revisable and open to reevaluation in the light of new evidence. Thus, if individuals routinely ignored the incoming data (e.g., behavior from their partners) and relied completely on their preexistent beliefs or lay theories (including relationship evaluations), then this would certainly show that love is blind (and that laypeople are arguably irrational). But as has been seen, the research evidence does not show this. Instead, it reveals in abundance that specific judgments and decisions in relationships are jointly determined by preexistent knowledge or lay theories and new information, and that relationship evaluations, beliefs, and the like are not set in concrete but change in response to new information.

If our model (see Figure 1) is on the right track, it implies that laypeople strive to maintain consistency among general relationship theories, local relationship theories, and specific judgments. This process constitutes an exemplary rational approach. Consider what happens in science, for example, if some new data are inconsistent with a standard model of the evolution or categorization of a particular star or volcano. If they are out of kilter, this will motivate either a reinterpretation of the data or a change in the local model or theory of the specific star or volcano, or perhaps at some point even lead to a modification of some general overarching theory. In a similar fashion, laypeople are motivated to keep the three levels of knowledge and theory (specific judgments, local theory, and general theory) in harmony in their social and relationship judgments (Fletcher, 1995).

Finally, our analysis of the evidence suggest that laypeople are not relationship Pollyannas but strategic optimists adjusting their local and general relationship theories in light of their goals, past experiences, the nature of the attributions involved, and contextual factors such as the stage of the relationship and the behavior of the partner. The way in which such factors work together can, and often do, push local relationship theories into negative territory, subsequently at times leading eventually to relationship dissolution.

Directions for Future Research

Our results, and analysis of the literature, suggest that the proposed model provides a useful tool for both integrating prior research findings and suggesting further avenues for research. However, one important direction for further research consists of drawing more on theories and approaches in fields outside romantic relationships. Such endeavors hold the promise of both learning more about relationships and contributing to our general understanding of bias and error in lay social judgments. We take an example of two approaches concerning the lay judgment of personality traits that have particular relevance but have seldom been applied to the relationship domain: the componential approach developed by Kenny et al. (2006) using the social relations model and Funder's (1995) realistic accuracy model.

Kenny's essentially methodological approach is based on the observation, originally developed by Cronbach (1955), that the factors producing accuracy in social judgments can be partitioned in several ways. For example, Stephen might produce accurate judgments of Mary because he is a good judge, or because Mary is easy to read, or because there is something about his specific relationship with Mary that produces high levels of accuracy. These kinds of effects can be pulled apart and examined best with round-robin designs, in which all participants are both perceivers and targets. Notably, this approach cannot be used for dyads in existing relationships. Nevertheless, features of the underlying methodology can be fruitfully applied to romantic relationships (for a detailed analysis, see Kenny et al., 2006), and the full design is applicable to speed-dating contexts, in which groups of individuals rate one another on various dimensions (although to our knowledge this design has not been used to assess accuracy in this context).

Funder's (1995) realistic accuracy approach also posits a version of a componential model. He argues that to assess personality traits accurately, perceivers must successfully negotiate four stages: (a) relevant cues must be displayed, (b) they must be accessible to the judge, (c) they must noticed and picked up by the judge, and (d) they must be correctly used to produce a judgment. The success of the judgment system is further thought to be a function of four distinct factors, which overlap those in the social relations model: (a) the ability of the judge, (b) the readability of the target, (c) the kind of trait being judged, and (d) the quantity and quality of information available to the judge.

A problem for accuracy research in any domain concerns the validity of the benchmark (the so-called criterion problem). The nature of the problem will vary according to the kind of judgment being made. For example, in the current meta-analysis, the benchmarks associated with predictions of relationship breakup or memories of prior relationship satisfaction possess a factual, objective quality. In contrast, measuring the "true" nature of personality or interaction traits, or fleeting thoughts and feelings, poses more serious difficulties. As noted, the most common benchmark used in the research consists of the self-perceived traits and mental activities of the target, but such self-perceptions will often be biased or flawed, or sometimes just plain wrong. Funder's (1995; Letzring et al., 2006) solution to this problem, which we concur with, is to view accuracy in construct validity terms. Thus, ideally researchers could use different benchmark criteria in the same study (such as self-judgments of the target, behavioral observations, clinical judgments, and so forth).

These ideas and theories developed in the study of lay personality attributions have not often been applied in the relationship domain, perhaps because the methodological requirements can be daunting. However, a few studies have revealed the gains that can be obtained from such approaches (see, e.g., Thomas & Fletcher, 2003).

The results of the meta-analyses and our related discussion point to many further questions and issues that require more research and theorizing. These include further work on the role of goals and motives for judgments, the question of the functionality of positive or negative biases, the extent to which excessive mean-level bias might cause problems for relationships, the role and nature of metacognition, the study of bias and error in the initial stages of mate selection, and how relationship cognition fits into broader evolutionary models. And, of course, understanding the role played by moderating and mediating variables in influencing the levels of mean-level bias and tracking accuracy is crucial to achieving a fuller understanding of relationship cognition in dyadic interaction. Finally, the study of the development of intimate relationships longitudinally from their earliest stages is rare (probably because it is so methodologically demanding). But such work promises much in terms of furthering understanding of the nature and functions of bias and accuracy of judgments in intimate relationships.

However, the results of the meta-analyses, and the research that we have reviewed more generally, reveal a plausible account for the functions and nature of mean-level bias in intimate relationships. In contrast, the functions of tracking accuracy in intimate relationships remain something of an enigma. It could be simply that tracking accuracy is developed to healthy levels early in the development of relationships to facilitate mating choices that maximize the chances of obtaining a good deal (relative to what the individual is offering) and avoiding disastrous choices. Alternatively, such tracking accuracy could operate as a backstop to prevent the blowout of mean-level bias from wreaking havoc when and if romantic love flourishes. If this is the case, then the combination of high levels of positive bias with poor tracking accuracy should be especially toxic over time in intimate relationships. Yet, no research evidence to our knowledge bears on this question. An important general direction for further research is to investigate the consequences for intimate relationships for different combinations of mean-level bias and tracking accuracy.

As noted already, we found no evidence from the meta-analyses that tracking accuracy was moderated by gender or relationship satisfaction (in stark contrast to mean-level bias). Nevertheless, the qualitative review of research above suggests that other moderators shown in Figure 1 do make a difference. For example, there is evidence that tracking accuracy improves when the goals of explanation, regulation, and prediction are made more accessible. Increased levels of accessibility, along with changes in tracking accuracy, also seem to be products of interactions between different moderators, such as context-driven stress and individual differences like self-esteem or attachment working models.

Conclusions

In the introduction we laid out a conundrum in the science of relationships; namely, love seems to be both blind and firmly rooted in the real world. The meta-analyses reported here go some way toward solving this puzzle, establishing that people can apparently be cheerleaders and seekers of the truth simultaneously; that is, they can be (and often are) both positively biased and accurate in judging their partners and relationships.

Our results and analyses are consistent with evolutionary approaches to romantic relationships. They support the existence of Darwinian sexual selection in humans, because people attain good levels of tracking accuracy on traits that are central to mate selection and retention. They are consistent with the argument that biases in relationship perception are adaptive, because such biases seem to be linked to differences in the costs and rewards involved in different outcomes. And they square elegantly with the proposal that romantic love is an evolved commitment device designed to lead men and women to substantially invest for long periods in one another and their offspring. The optimistic spin that individuals put on their relationship judgments is, on this account, a product of ancient, evolved adaptations.

Our results, and review of the literature, also support a standard social psychological approach in which laypeople strive to achieve balance between their relationship theories and their judgments, and in which (relationship-level) affect, behavior, and cognition are profoundly interdependent. Lay intimate relationship theories and judgments are typically not castles in the air, remote from partners and the consequences. Rather, such theories and judgments develop in small and intense groups (dyads) in which participants' fates are intertwined. In intimate relationships, reality and illusion go hand in hand in the furtherance of goals that have a long evolutionary history and where the outcomes, for good or ill, have profound personal consequences.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Acitelli, L. K., Kenny, D. A., & Weiner, D. (2001). The importance of similarity and understanding of partners' marital ideals to relationship satisfaction. *Personal Relationships, 8,* 167–185.

*Agnew, C. R., Loving, T. J., & Drigotas, S. M. (2001). Substituting the forest for the trees: Social networks and the prediction of romantic relationship state and fate. *Journal of Personality and Social Psychology, 81,* 1042–1057.

*Andrews, P. W., Gangestad, S. W., Miller, G. F., Haselton, M. G., Thornhill, R., & Neale, M. C. (2008). Sex differences in detecting sexual

infidelity: Results of a maximum likelihood method for analyzing the sensitivity of sex differences to underreporting. *Human Nature, 19,* 347–373.

Archer, J. (1999). Assessment of the reliability of the Conflict Tactics Scales: A meta-analytic review. *Journal of Interpersonal Violence, 14,* 1263–1289.

Baldwin, M. W. (Ed.). (2006). *Interpersonal cognition.* New York, NY: Guilford Press.

*Barelds-Dijkstra, P., & Barelds, D. P. H. (2008). Positive illusions about one's partner's physical attractiveness. *Body Image, 5,* 99–108.

Beer, A., & Watson, D. (2008). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment, 90,* 250–260.

*Bell, M. E., Cattaneo, L. B., Goodman, L. A., & Dutton, M. A. (2008). Assessing the risk of future psychological abuse: Predicting the accuracy of battered women's predictions. *Journal of Family Violence, 23,* 69–70.

Bernieri, F. J., Zuckerman, M., Koestner, R., & Rosenthal, R. (1994). Measuring person perception accuracy: Another look at self-other agreement. *Personality and Social Psychology Bulletin, 20,* 367–378.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley.

*Boyes, A. D., & Fletcher, G. J. O. (2007). Metaperceptions of bias in intimate relationships. *Journal of Personality and Social Psychology, 92,* 286–306.

*Calhoun, P. S., Beckham, C. C., Feldman, M. E., Barefoot, J. C., Haney, T., & Bosworth, H. B. (2002). Partners' ratings of combat veterans' anger. *Journal of Traumatic Stress, 15,* 133–136.

Campbell, L., Lackenbauer, S. D., & Muise, A. (2006). When is being known or adored by romantic partners most beneficial? Self-perceptions, relationship length, and responses to partner's verifying and enhancing appraisals. *Personality and Social Psychology Bulletin, 32,* 1283–1294.

Campbell, L., Simpson, J. A., Kashy, D. A., & Fletcher, G. J. O. (2001). Ideal standards, the self, and flexibility of ideals in close relationships. *Personality and Social Psychology Bulletin, 27,* 447–462.

*Capaldi, D. M., & Owen, L. D. (2001). Physical aggression in a community sample of at-risk young couples: Gender comparisons for high frequency, injury, and fear. *Journal of Family Psychology, 15,* 425–440.

*Capaldi, D. M., Shortt, J. W., & Crosby, L. (2003). Physical and psychological aggression in at-risk young couples: Stability and change in young adulthood. *Merrill-Palmer Quarterly, 49,* 1–27.

*Cattaneo, L. B., Bell, M. E., Goodman, L. A., & Dutton, M. A. (2007). Intimate partner violence victims' accuracy in assessing their risk of re-abuse. *Journal of Family Violence, 22,* 429–440.

*Cattaneo, L. B., & Goodman, L. A. (2003). Victim-reported risk factors for continued abusive behavior: Assessing the dangerousness of arrested batterers. *Journal of Community Psychology, 31,* 349–369.

Chen, S., Boucher, H. C., & Tapias, M. P. (2006). The relational self revealed: Integrative conceptualization and implications for interpersonal life. *Psychological Bulletin, 132,* 151–179.

*Clements, K., Holtzworth-Munroe, A., Schweinle, W., & Ickes, W. (2007). Empathic accuracy of intimate partners in violent versus nonviolent relationships. *Personal Relationships, 14,* 369–388.

Collins, N. L., & Allard, L. M. (2001). Cognitive representations of attachment: The content and function of working models. In G. J. O. Fletcher & M. S. Clark (Eds.), *Blackwell handbook of social psychology: Interpersonal processes* (pp. 60–85). Malden, MA: Blackwell.

*Collins, N. L., & Feeney, B. C. (2000). A safe haven: An attachment theory perspective on support seeking and caregiving in intimate relationships. *Interpersonal Relations and Group Processes, 78,* 1053–1076.

*Collins, N. L., & Feeney, B. C. (2004). Working models of attachment shape perceptions of social support: Evidence from experimental and observational studies. *Journal of Personality and Social Psychology, 87,* 363–383.

Collins, N. L., & Read, S. J. (1994). Cognitive representations of attachment: The structure and function of working models. In K. Bartholomew & D. Perlman (Eds.), *Advances in Personal Relationships: Vol. 5. Attachment processes in adulthood* (pp. 53–90). London, England: Kingsley.

Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin, 116,* 3–20.

*Corsini, R. J. (1956). Understanding and similarity in marriage. *Journal of Abnormal and Social Psychology, 52,* 327–332.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52,* 177–193.

*Cui, M., Lorenz, F. O., Conger, R. D., Melby, J. N., & Bryant, C. M. (2005). Observer, self-, and partner reports of hostile behaviors in romantic relationships. *Journal of Marriage and Family, 67,* 1169–1181.

De La Ronde, C., & Swann, W. B., Jr. (1998). Partner verification: Restoring shattered images of our intimates. *Journal of Personality and Social Psychology, 75,* 374–382.

*Drigotas, S. M., Rusbult, C. E., Wieselquist, J., & Whitton, S. W. (1999). Close partner as sculptor of the ideal self: Behavioral affirmation and the Michelangelo phenomenon. *Journal of Personality and Social Psychology, 77,* 293–323.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177.

*Dunn, E. W., Biesanz, J. C., Human, L. J., & Finn, S. (2007). Misunderstanding the affective consequences of everyday social interactions: The hidden benefits of putting one's best face forward. *Journal of Personality and Social Psychology, 92,* 990–1005.

Dunning, D., Kunda, Z., & Murray, S. L. (1999). What the commentators motivated us to think about. *Psychological Inquiry, 10,* 79–82.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89–98.

Dymond, R. (1954). Interpersonal perception and marital happiness. *Canadian Journal of Psychology, 8,* 164–171.

Eastwick, P. W., & Finkel, E. J. (2008). Sex differences in mate preferences revisited: Do people know what they initially desire in a romantic partner? *Journal of Personality and Social Psychology, 94,* 245–264.

*Eastwick, P. W., Finkel, E. K., Krishnamurti, T., & Loewenstein, G. (2008). Mispredicting distress following romantic breakup: Revealing the time course of the affective forecasting error. *Journal of Experimental Social Psychology, 44,* 800–807.

*Edwards, A. L., & Klockars, A. J. (1981). Significant others and self-evaluation: Relationships between perceived and actual evaluations. *Personality and Social Psychology Bulletin, 7,* 244–251.

*Epley, N., & Dunning, D. (2006). The mixed blessings of self-knowledge in behavioral prediction: Enhanced discrimination but exacerbated bias. *Personality and Social Psychology Bulletin, 32,* 641–655.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Fiedler, K., & Wänke, M. (2009). The cognitive–ecological approach to rationality in social psychology. *Social Cognition, 27,* 699–732.

Fisher, H. E., Aron, A., & Brown, L. L. (2006). Romantic love: A mammalian brain system for mate choice. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361,* 2173–2186.

Fitness, J., & Peterson, J. (2008). Punishment and forgiveness in close relationships: An evolutionary, social-psychological analysis. In J. P. Forgas & J. Fitness (Eds.), *Social relationships: Cognitive, affective, and motivational processes* (pp. 255–270). Hove, England: Psychology Press.

Fletcher, G. J. O. (1995). *The scientific credibility of folk psychology.* Hillsdale, NJ: Erlbaum.

Fletcher, G. J. O. (2002). *The new science of intimate relationships.* Malden, MA: Blackwell.

Fletcher, G. J. O., & Boyes, A. D. (2008). Is love blind? Reality and illusion in intimate relationships. In J. P. Forgas & J. Fitness (Eds.), *Social relationships: Cognitive, affective, and motivational processes* (pp. 101–114). Hove, England: Psychology Press.

Fletcher, G. J. O., & Fincham, F. D. (Eds.). (1991). *Cognition in close relationships.* Hillsdale, NJ: Erlbaum.

Fletcher, G. J. O., & Fitness, J. (Eds.). (1996). *Knowledge structures in close relationships: A social psychological approach.* Mahwah, NJ: Erlbaum.

Fletcher, G. J. O., Overall, N., & Friesen, M. (2006). Social cognition in intimate relationships. In A. L. Vangelisti & D. Perlman (Eds.), *The Cambridge handbook of personal relationships* (pp. 353–368). New York, NY: Cambridge University Press.

Fletcher, G. J. O., Reeder, G. D., & Bull, V. (1990). Bias and accuracy in attitude attribution: The role of attributional complexity. *Journal of Experimental Social Psychology, 26,* 275–288.

Fletcher, G. J. O., Simpson, J. A., & Boyes, A. D. (2006). Accuracy and bias in romantic relationships: An evolutionary and social psychological analysis. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 189–209). New York, NY: Psychology Press.

Fletcher, G. J. O., Simpson, J. A., & Thomas, G. (2000). Ideals, perceptions, and evaluations in early relationship development. *Journal of Personality and Social Psychology, 79,* 933–940.

Fletcher, G. J. O., Simpson, J. A., Thomas, G., & Giles, L. (1999). Ideals in intimate relationships. *Journal of Personality and Social Psychology, 76,* 72–89.

Fletcher, G. J. O., & Thomas, G. (1996). Close relationship lay theories: Their structure and function. In G. J. O. Fletcher & J. Fitness (Eds.), *Knowledge structures in close relationships* (pp. 3–24). Mahwah, NJ: Erlbaum.

Fletcher, G. J. O., & Thomas, G. (2000). Behavior and on-line cognition in marital interaction. *Personal Relationships, 7,* 111–130.

*Foltz, C., Morse, J. Q., & Barber, J. P. (1999). Self and observer reports of interpersonal problems in couples. *Journal of Clinical Psychology, 55,* 27–37.

*Försterling, F., Schuster, B., & Morgenstern, M. (2005). Perspectives of causal attributions: A study of couples and depression. *Journal of Social and Clinical Psychology, 24,* 793–810.

Fowers, B. J., Lyons, E., Montel, K. H., & Shaked, N. (2001). Positive illusions about marriage among married and single individuals. *Journal of Family Psychology, 15,* 95–109.

*Friesen, M. D., Fletcher, G. J. O., & Overall, N. C. (2005). A dyadic assessment of forgiveness in intimate relationships. *Personal Relationships, 12,* 61–77.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102,* 652–670.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55,* 149–158.

Funder, D. C., & Colvin, C. R. (1997). Congruence of others' and self-judgments of personality. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 617–647). San Diego, CA: Academic Press.

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52,* 409–418.

Gable, S. L., & Poore, J. (2008). Which thoughts count? Algorithms for evaluating satisfaction in relationships. *Psychological Science, 19,* 1030–1036.

*Gagné, F. M., & Lydon, J. E. (2001). Mind-set and close relationships: When bias leads to (in)accurate predictions. *Journal of Personality and Social Psychology, 81,* 85–96.

*Gagné, F. M., & Lydon, J. E. (2003). Identification and the commitment shift: Accounting for gender differences in relationship illusions. *Personality and Social Psychology Bulletin, 29,* 907–919.

Gagné, F. M., & Lydon, J. E. (2004). Bias and accuracy in close relationships: An integrative review. *Personality and Social Psychology Review, 8,* 322–338.

*Gauthier, N., Thibaut, P., & Sullivan, M. J. L. (2008). Individual and relational correlates of pain-related empathic accuracy in spouses of chronic pain patients. *Clinical Journal of Pain, 24,* 669–677.

Gill, M. J., & Swann, W. B., Jr. (2004). What it means to know someone: A matter of pragmatics. *Journal of Personality and Social Psychology, 86,* 405–418.

Gillath, O., Mikulincer, M., Fitzsimons, G. M., Shaver, P. R., Schachner, D. A., & Bargh, J. A. (2006). Automatic activation of attachment-related goals. *Personality and Social Psychology Bulletin, 32,* 1375–1388.

Gollwitzer, P. M. (2003). Why we thought that action mind-sets affect illusions of control. *Psychological Inquiry, 14,* 261–269.

*Gomà-i-Freixanet, M. (1997). Consensual validity of the EPQ: Self-reports and spouse-reports. *European Journal of Psychological Assessment, 13,* 179–185.

Gonzaga, G. C., & Haselton, M. G. (2008). The evolution of love and long-term bonds. In J. P. Forgas & J. Fitness (Eds.), *Social relationships: Cognitive, affective, and motivational processes* (pp. 39–54). Hove, England: Psychology Press.

*Goodwin, S. A., Fiske, S. T., Rosen, L. D., & Rosenthal, A. M. (2002). The eye of the beholder: Romantic goals and impression biases. *Journal of Experimental Social Psychology, 38,* 232–241.

Griffin, D., Murray, S., Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships, 6,* 505–518.

Haig, B. D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *American Journal of Psychology, 122,* 219–234.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85,* 845–857.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78,* 81–91.

Hassin, R. R. (2008). Cognitive motivation and motivated cognition: A preface. *Social Cognition, 26,* 495.

*Heckert, D. A., & Gondolf, E. W. (2004). Battered women's perceptions of risk versus risk factors and instruments in predicting repeat reassault. *Journal of Interpersonal Violence, 19,* 778–800.

Heider, F. (1958). *The psychology of interpersonal relations.* New York, NY: Wiley.

*Holzworth-Munroe, A., Rehman, U., & Herron, K. (2000). General and spouse-specific anger and hostility in subtypes of martially violent men and nonviolent men. *Behavior Therapy, 31,* 603–630.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61,* 587–610.

Ickes, W., Stinton, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology, 59,* 730–742.

*Julian, T. W., McKenry, P. C., Gavazzi, S. M., & Law, J. C. (1999). Test of family of origin structural models of male verbal and physical aggression. *Journal of Family Issues, 20,* 397–423.

*Kahn, M. (1970). Non-verbal communication and marital satisfaction. *Family Processes, 9,* 449–456.

Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, methods, and research. *Psychological Bulletin, 118,* 3–34.

*Karney, B. R., & Coombs, R. H. (2000). Memory bias in long-term close

relationships: Consistency or improvement? *Personality and Social Psychology Bulletin, 26,* 959–970.

*Karney, B. R., & Frye, N. E. (2002). "But we've been getting better lately": Comparing prospective and retrospective views of relationship development. *Journal of Personality and Social Psychology, 82,* 222–238.

Kelley, H. H. (1979). *Personal relationships: Their structures and processes.* New York, NY: Wiley.

*Kenny, D. A., & Acitelli, L. K. (2001). Accuracy and bias in the perception of the partner in a close relationship. *Journal of Personality and Social Psychology, 80,* 439–448.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102,* 390–402.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis.* New York, NY: Guilford Press.

Kenny, D. A., & Winquist, L. A. (2001). The measurement of interpersonal sensitivity: Consideration of design, components, and units of analysis. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 265–302). Mahwah, NJ: Erlbaum.

*Khawaja, M., & Tewtel-Salem, M. (2004). Agreement between husband and wife reports of domestic violence: Evidence from poor refugee communities in Lebanon. *International Journal of Epidemiology, 33,* 526–533.

*Kilpatrick, S. D., Bissonnette, V. L., & Rusbult, C. E. (2002). Empathic accuracy and accommodative behavior among newly married couples. *Personal Relationships, 9,* 369–393.

*Kirchler, E. (1999). Unbelievable similarity: Accuracy in spouses' reports on their partners' tactics to influence joint economic decisions. *Applied Psychology: An International Review, 48,* 329–348.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228.

Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin, 27,* 720–730.

Knee, C. R., Patrick, H., & Lonsbary, C. (2003). Implicit theories of relationships: Orientations toward evaluation and cultivation. *Personality and Social Psychology Review, 7,* 41–55.

*Koerner, A. F., & Fitzpatrick, M. A. (2002). Nonverbal communication and marital adjustment and satisfaction: The role of decoding relationship relevant and relationship irrelevant affect. *Communication Monographs, 69,* 33–51.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences, 27,* 313–376.

*Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53,* 636–647.

Lackenbauer, S. D., Campbell, L., Fletcher, G. J. O., & Rubin, H. (in press). Seeing clearly through rose-colored glasses: The unique and combined benefits of being perceived accurately and in a positively biased manner by romantic partners. *Personal Relationships.*

*Lawrence, E., & Bradbury, T. N. (2001). Physical aggression and marital dysfunction: A longitudinal analysis. *Journal of Family Psychology, 15,* 135–154.

*Lawrence, E., & Bradbury, T. N. (2007). Trajectories of change in physical aggression and marital satisfaction. *Journal of Family Psychology, 21,* 236–247.

*Lemay, E. P., Jr., & Clark, M. S. (2008). How the head liberates the heart: Projection of communal responsiveness guides relationship promotion. *Journal of Personality and Social Psychology, 4,* 647–671.

*Lemay, E. P., Jr., Pruchno, R. A., & Feild, L. (2006). Accuracy and bias in perceptions of spouses' life-sustaining medical treatment preferences. *Journal of Applied Social Psychology, 36,* 2337–2361.

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology, 91,* 111–123.

*Lorenz, F. O., Melby, J. N., Conger, R. D., & Xu, X. (2007). The effects of context on the correspondence between observational ratings and questionnaire reports of hostile behavior: A multitrait, multimethod approach. *Journal of Family Psychology, 21,* 498–509.

*Loving, T. J. (2006). Predicting dating relationship fate with insiders' and outsiders' perspectives: Who and what is asked matters. *Personal Relationships, 13,* 349–362.

*MacDonald, T. K., & Ross, M. (1999). Assessing the accuracy of predictions about dating relationships: How and why do lovers' predictions differ from those made by observers? *Personality and Social Psychology Bulletin, 25,* 1417–1429.

Marcus, D. K., & Miller, R. S. (2003). Sex differences in judgments of physical attractiveness: A social relations analysis. *Personality and Social Psychology Bulletin, 29,* 325–335.

*Marshall, A. D., & Holtzworth-Munroe, A. (2002). Varying forms of husband sexual aggression: Predictors and subgroup differences. *Journal of Family Psychology, 16,* 286–296.

Mast, M. S., & Hall, J. A. (2004). Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior, 28,* 145–165.

*McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology, 43,* 293–303.

*McCrae, R. R. (1993). Agreement of personality profiles across observers. *Multivariate Behavioral Research, 28,* 25–40.

*McCrae, R. R., Stone, S. V., Fagan, P. J., & Costa, P. T., Jr. (1998). Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality, 66,* 285–312.

*McFarland, C., & Ross, M. (1987). The relation between current impressions and memories of self and dating partners. *Personality and Social Psychology Bulletin, 13,* 228–238.

McNulty, J. K., & Karney, B. R. (2004). Positive expectations in the early years of marriage: Should couples expect the best or brace for the worst? *Journal of Personality and Social Psychology, 86,* 729–743.

McNulty, J. K., O'Mara, E. M., & Karney, B. R. (2008). Benevolent cognitions as a strategy of relationship maintenance: "Don't sweat the small stuff". . . . But it is not all small stuff. *Journal of Personality and Social Psychology, 94,* 631–646.

Mikulincer, M., & Arad, D. (1999). Attachment working models and cognitive openness in close relationships: A test of chronic and temporary accessibility effects. *Journal of Personality and Social Psychology, 77,* 710–725.

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships, 25,* 889–922.

Murray, S. L. (2001). Seeking a sense of conviction: Motivated cognition in close relationships. In G. J. O. Fletcher & M. S. Clark (Eds.), *Blackwell handbook of social psychology: Interpersonal processes* (pp. 107–126). Malden, MA: Blackwell.

*Murray, S. L., Bellavia, G. M., Rose, P., & Griffin, D. W. (2003). Once hurt, twice hurtful: How perceived regard regulates daily marital interactions. *Journal of Personality and Social Psychology, 84,* 126–147.

Murray, S. L., Griffin, D. W., Rose, P., & Bellavia, G. M. (2003). Calibrating the sociometer: The relational contingencies of self-esteem. *Journal of Personality and Social Psychology, 85,* 63–84.

Murray, S. L., Griffin, D. W., Rose, P., & Bellavia, G. M. (2006). For better or worse? Self-esteem and the contingencies of acceptance in marriage. *Personality and Social Psychology Bulletin, 32,* 866–880.

Murray, S. L., & Holmes, J. G. (1993). Seeing virtues in faults: Negativity and the transformation of interpersonal narratives in close relationships. *Journal of Personality and Social Psychology, 65,* 707–722.

Murray, S. L., & Holmes, J. G. (2009). The architecture of interdependent minds: A motivation-management theory of mutual responsiveness. *Psychological Review, 116,* 908–928.

*Murray, S. L., Holmes, J. G., Bellavia, G., Griffin, D. W., & Dolderman, D. (2002). Kindred spirits? The benefits of egocentrism in close relationships. *Journal of Personality and Social Psychology, 82,* 563–581.

Murray, S. L., Holmes, J. G., & Collins, N. L. (2006). Optimizing assurance: The risk regulation system in relationships. *Psychological Bulletin, 132,* 641–666.

*Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996a). The benefits of positive illusions: Idealization and the construction of satisfaction in close relationships. *Journal of Personality and Social Psychology, 70,* 79–98.

*Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996b). The self-fulfilling nature of positive illusions in romantic relationships: Love is not blind, but prescient. *Journal of Personality and Social Psychology, 71,* 1155–1180.

Murray, S. L., Holmes, J. G., & Griffin, D. W. (2000). Self-esteem and the quest for felt security: How perceived regard regulates attachment processes. *Journal of Personality and Social Psychology, 78,* 478–498.

*Murray, S. L., Holmes, J. G., Griffin, D. W., Bellavia, G., & Rose, P. (2001). The mismeasure of love: How self-doubt contaminates relationship beliefs. *Personality and Social Psychology Bulletin, 27,* 423–436.

*Mutén, E. (1991). Self-reports, spouse ratings, and psychophysiological assessment in a behavioral medicine program: An application of the five-factor model. *Journal of Personality Assessment, 57,* 449–464.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

*Noller, P. (1980). Misunderstandings in marital communication: A study of couples' nonverbal communication. *Journal of Personality and Social Psychology, 39,* 1135–1148.

*Noller, P., & Feeny, J. A. (1994). Relationship satisfaction, attachment, and nonverbal accuracy in early marriage. *Journal of Nonverbal Behavior, 18,* 199–221.

*Oishi, S., & Sullivan, H. W. (2006). The predictive value of daily vs. retrospective well-being judgments in relationship stability. *Journal of Experimental Social Psychology, 42,* 460–470.

*O'Leary, K. D., & Williams, M. C. (2006). Agreement about acts of aggression in marriage. *Journal of Family Psychology, 20,* 656–662.

Overall, N. C., & Fletcher, G. J. O. (in press). Perceiving regulation from intimate partners: Reflected appraisal and self-regulation processes in close relationships. *Personal Relationships.*

Overall, N. C., Fletcher, G. J. O., & Friesen, M. D. (2003). Mapping the intimate relationship mind: Comparisons between three models of attachment representations. *Personality and Social Psychology Bulletin, 29,* 1479–1493.

*Overall, N. C., Fletcher, G. J. O., & Simpson, J. A. (2006). Regulation processes in intimate relationships: The role of ideal standards. *Journal of Personality and Social Psychology, 91,* 662–685.

Overall, N. C., Fletcher, G. J. O., Simpson, J. A., & Sibley, C. G. (2009). Regulating partners in intimate relationships: The costs and benefits of different communication strategies. *Journal of Personality and Social Psychology, 96,* 620–639.

*Panuzio, J., O'Farrell, T. J., Marshall, A. D., Murphy, C. M., Murphy, M., & Taft, C. T. (2006). Intimate partner aggression reporting concordance and correlates of agreement among men with alcohol use disorders and their female partners. *Assessment, 13,* 266–279.

*Perry, A. R., & Fromuth, M. E. (2005). Courtship violence using couple data: Characteristics and perceptions. *Journal of Interpersonal Violence, 20,* 1078–1095.

Pierce, T., & Lydon, J. E. (2001). Global and specific relational models in the experience of social interactions. *Journal of Personality and Social Psychology, 80,* 613–631.

Place, S. K., Todd, P. M., Penke, L., & Asendorpf, J. B. (2009). The ability to judge the romantic interest of others. *Psychological Science, 20,* 22–26.

Reis, H. T., Clark, M. S., & Holmes, J. G. (2004). Perceived partner responsiveness as an organizing construct in the study of intimacy and closeness. In D. J. Mashek & A. P. Aron (Eds.), *Handbook of closeness and intimacy* (pp. 201–225). Mahwah, NJ: Erlbaum.

Reis, H. T., & Patrick, B. C. (1996). Attachment and intimacy: Component processes. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 367–389). New York, NY: Guilford Press.

Reis, H. T., & Shaver, P. R. (1988). Intimacy as an interpersonal process. In S. Duck (Ed.), *Handbook of personal relationships: Theory, research, and interventions* (pp. 367–389). Chichester, England: Wiley.

*Rhoades, G. K., & Stocker, C. M. (2006). Can spouses provide knowledge of each other's communication patterns? A study of self-reports, spouses' reports, and observational coding. *Family Process, 45,* 499–511.

*Ro, E., & Lawrence, E. (2007). Comparing three measures of psychological aggression: Psychometric properties and differentiation from negative communication. *Journal of Family Violence, 22,* 575–586.

*Robinson, E. A., & Price, M. G. (1980). Pleasurable behavior in marital interaction: An observational study. *Journal of Consulting and Clinical Psychology, 48,* 117–118.

Rusbult, C. E., Van Lange, P. A. M., Wildschut, T., Yovetich, N. A., & Verette, J. (2000). Perceived superiority in close relationships: Why it exists and persists. *Journal of Personality and Social Psychology, 79,* 521–545.

*Sabatelli, R. M., Buck, R., & Dreyer, A. (1982). Nonverbal communication accuracy in married couples: Relationship with marital complaints. *Journal of Personality and Social Psychology, 43,* 1088–1097.

*Saffrey, C., Bartholomew, K., Scharfe, E., Henderson, A. J. Z., & Koopman, R. (2003). Self- and partner-perceptions of interpersonal problems and relationship functioning. *Journal of Social and Personal Relationships, 20,* 117–139.

*Sagrestano, L. M., Heavey, C. L., & Christensen, A. (1999). Perceived power and physical violence in marital conflict. *Journal of Social Issues, 55,* 65–79.

*Sanderson, C. A., & Cantor, N. (2001). The association of intimacy goals and marital satisfaction: A test of four mediational hypotheses. *Personality and Social Psychology Bulletin, 27,* 1567–1577.

*Sanderson, C. A., & Evans, S. M. (2001). Seeing one's partner through intimacy-colored glasses: An examination of the processes underlying the intimacy goals–relationship satisfaction link. *Personality and Social Psychology Bulletin, 27,* 463–473.

*Schul, Y., & Vinokur, A. D. (2000). Projection in person perception among spouses as a function of the similarity in their shared experiences. *Personality and Social Psychology Bulletin, 26,* 987–1001.

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology, 84,* 60–79.

*Senécal, S., Murad, N., & Hess, U. (2003). Do you know what I feel? Partners' predictions and judgments of each other's emotional reactions to emotion-eliciting situations. *Sex Roles, 48,* 21–37.

*Sillars, A. L., Pike, G. R., Jones, T. S., & Murphy, M. A. (1984). Communication and understanding in marriage. *Human Communication Research, 10,* 317–350.

*Sillars, A. L., Weisberg, J., Burggraf, C. S., & Zietlow, P. H. (1990). Communication and understanding revisited: Married couples' understanding and recall of conversations. *Communication Research, 17,* 500–522.

Simpson, J. A., Ickes, W., & Blackstone, T. (1995). When the head protects the heart: Empathic accuracy in dating relationships. *Journal of Personality and Social Psychology, 69,* 629–641.

Simpson, J. A., Ickes, W., & Grich, J. (1999). When accuracy hurts: Reactions of anxious–ambivalent dating partners to a relationship-

threatening situation. *Journal of Personality and Social Psychology, 76,* 754–769.

*Simpson, J. A., Oriña, M. M., & Ickes, W. (2003). When accuracy hurts, and when it helps: A test of the empathic accuracy model in marital interactions. *Journal of Personality and Social Psychology, 85,* 881–893.

*Simpson, L. E., & Christensen, A. (2005). Spousal agreement regarding relationship aggression on the Conflict Tactics Scale–2. *Psychological Assessment, 17,* 423–432.

*Smith, D. A., & Peterson, K. M. (2008). Overperception of spousal criticism in dysphoria and marital discord. *Behavior Therapy, 39,* 300–312.

*Smith, L., Heaven, P. C. L., & Ciarrochi, J. (2008). Trait emotional intelligence, conflict communication patterns, and relationship satisfaction. *Personality and Individual Differences, 44,* 1314–1325.

*South, S. J., Bose, S., & Trent, K. (2004). Anticipating divorce: Spousal agreement, predictive accuracy, and effects on labor supply and fertility. *Journal of Divorce & Remarriage, 40,* 1–22.

*Sprecher, S. (1999). "I love you more today than yesterday": Romantic partners' perceptions of changes in love and related affect over time. *Journal of Personality and Social Psychology, 76,* 46–53.

Strachman, A., & Gable, S. L. (2006). Approach and avoidance relationship commitment. *Motivation and Emotion, 30,* 117–126.

Swann, W. B., Jr., Bosson, J. K., & Pelham, B. W. (2002). Different partners, different selves: Strategic verification of circumscribed identities. *Personality and Social Psychology Bulletin, 28,* 1215–1228.

Swann, W. B., Jr., De La Ronde, C., & Hixon, J. G. (1994). Authenticity and positivity strivings in marriage and courtship. *Journal of Personality and Social Psychology, 66,* 857–869.

*Swann, W. B., Jr., & Gill, M. J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology, 73,* 747–757.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193–210.

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups.* New York, NY: Wiley.

*Thomas, G., & Fletcher, G. J. O. (2003). Mind-reading accuracy in intimate relationships: Assessing the roles of the relationship, the target, and the judge. *Journal of Personality and Social Psychology, 85,* 1079–1094.

*Thomas, G., Fletcher, G. J. O., & Lange, C. (1997). On-line empathic accuracy in marital interaction. *Journal of Personality and Social Psychology, 72,* 839–850.

*Tiggle, R. B., Peters, M. D., Kelley, H. H., & Vincent, J. (1982). Correlational and discrepancy indices of understanding and their relation to marital satisfaction. *Journal of Marriage & Family, 44,* 209–215.

Todd, P. M., Penke, L., Fasolo, B., & Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proceedings of the National Academy of Sciences, USA, 104,* 15011–15016.

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man: 1871–1971* (pp. 136–179). Chicago, IL: Aldine.

*Tucker, J. S., & Anders, S. L. (1999). Attachment style, interpersonal perception accuracy, and relationship satisfaction in dating couples. *Personality and Social Psychology Bulletin, 25,* 403–412.

*Van Lange, P. A. M., Rusbult, C. E., Drigotas, S. M., Arriaga, X. B., Witcher, B. S., & Cox, C. L. (1997). Willingness to sacrifice in close relationships. *Journal of Personality and Social Psychology, 72,* 1373–1395.

*Waldinger, R. J., & Shulz, M. S. (2006). Linking hearts and minds in couple interactions: Intentions, attributions, and overriding sentiments. *Journal of Family Psychology, 20,* 494–504.

*Watson, D., Hubbard, B., & Wiese, D. (2000). Self–other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology, 78,* 546–558.

*Watson, D., & Humrichouse, J. (2006). Personality development in emerging adulthood: Integrating evidence from self-ratings and spouse ratings. *Journal of Personality and Social Psychology, 91,* 959–974.

*Weigel, D. J. (2008). A dyadic assessment of how couples indicate their commitment to each other. *Personal Relationships, 15,* 17–39.

*Weisz, A. N., Tolman, R. M., & Saunders, D. G. (2000). Assessing the risk of severe domestic violence: The importance of survivors' predictions. *Journal of Interpersonal Violence, 15,* 75–90.

*Werner, P. D., Green, R.-J., Greenberg, J., Browne, T. L., & McKenna, T. E. (2001). Beyond enmeshment: Evidence for the independence of intrusiveness and closeness–caregiving in married couples. *Journal of Marital and Family Therapy, 27,* 459–471.

*Wilhelm, P., & Perrez, M. (2004). How is my partner feeling in different daily-life settings? Accuracy of spouses' judgements about their partner's feelings at work and at home. *Social Indicators Research, 67,* 183–246.

Received June 21, 2009
Revision received March 18, 2010
Accepted March 23, 2010 ■