Perceiving Personality Through the Grapevine: A Network Approach to Reputations

Cory K. Costello & Sanjay Srivastava
University of Oregon

Working Paper – subject to revision
08/09/2019

Correspondence should be sent to:
Cory K. Costello
Psychology Department
1227 University of Oregon
Eugene, OR 97403
email: ccostell@uoregon.edu

Abstract

Reputations are critical in human social life: they allow people to share and act on information about one another, even when they have never met. Reputations can be conceptualized as information about a target person that is stored in networks of perceivers and transmitted through either direct interaction or hearsay. We present a novel paradigm that integrates the network approach with concepts and methods from interpersonal perception research. We apply that paradigm to study the relative valence, consensus, accuracy, and consequences of personality trait information in hearsay-based reputations. In two preregistered studies (N = 260 and 369), we created naturalistic micronetworks in the lab in which participants interacted and got to know one another, then later described each other to naïve third parties. Across studies, hearsay-based reputations were less positive than direct reputations and self-perceptions. Hearsay reputations demonstrated appreciable consensus (agreement) with direct reputations but only modest accuracy, suggesting that they provide a robust but substantially inaccurate signal. Neither hearsay consensus nor accuracy were moderated by individual differences in extraversion, empathy, or perspective-taking. In Study 2 we experimentally manipulated participants' impression-formation goals, but this had no detectable impact on hearsay consensus or accuracy. Hearsay reputations were consequential, affecting the extent to which perceivers thought targets would be good leaders or friends. These results provide initial insights into reputation networks and suggest several important future directions for the network approach to reputations. We also present open materials and data analysis software for others to extend the reputations-as-networks approach.

Keywords: Reputation; Personality; Interpersonal Perception; Gossip

Perceiving People Through the Grapevine: A Network Approach to Reputations

Familiar phrases like "your reputation precedes you" capture a simple truth about our social world: reputations extend beyond people who have met one another, communicated further through the processes of gossip and hearsay (Craik, 2009). Sharing reputational knowledge is thought to at least partly underpin the unprecedented scale and cooperativeness of human social groups, by both expanding the volume of our social knowledge and allowing communities to act on shared knowledge about an individual (Dunbar, 2004; Feinberg, Willer, & Schultz, 2014; Wu et al., 2015). Indeed, the desire and ability to share reputational knowledge is likely one factor behind the rich personality lexicons found in natural languages, which have served as a databases for identifying structural models of personality like the Big 5 and Big 6 (Saucier & Goldberg, 1996; Saucier & Srivastava, 2015; Wood, 2015). Some theories even suggest that trait terms are primarily for describing others (Buss & Craik, 1983; Hogan, 1994; Srivastava, 2010; Saucier & Srivastava, 2015; Wood, Tov, & Costello, 2015). Documenting how people describe each other – how reputational knowledge is shared – thus has far-ranging theoretical implications.

We present here a *network approach* to the study of reputation. We extend the methodological and analytic tools of Kenny's (1994) interpersonal perception paradigm to study the questions raised by Craik's (2009) network interpretation of reputations, ultimately providing a paradigm to study the flow of reputational information. Previous work on interpersonal perception has limited its attention to reputations formed by individuals perceiving one another, an approach that we refer to as the *dyadic approach* to reputation because of its focus on the target-perceiver (or self-other) dyad (e.g., Funder, 1995; Kenny, 1994, 2004; McAbee Connelly, 2016; Vazire, 2010). In the network approach, this process – the transmission of information

from target to perceiver – is embedded in a larger network, where that perceiver also transmits information about the target to other perceivers who have not met the target, and so on. In the next section, we introduce the concept of reputation networks and provide an overview of the network approach to reputation, building on the dyadic approach.

Reputation networks

In the most general terms, network analysis concerns the relations (called *edges*) between entities (called *nodes*), collectively referred to as a *network* (Kadushin, 2012). In Craik's (2009) network approach to reputations, the nodes (*T*s and *P*s) represent perceivers (*P*) who know or know of some target (*T*). The edges represent the flow of information about T between one or more nodes (from T to Ps and between Ps). Reputation networks are thus egocentric networks where T is the *ego,* Ps are the *alters,* and edges represent information transmitted about T. In short, a reputational network is made up of nodes (people) that store and transmit information about a target person. Figure 1 depicts a simple, hypothetical reputational network. Some nodes are directly connected to T, Ts first-degree connections, which we refer to as *first-degree perceivers* (*P1s*). Other nodes will not be directly connected to T, but *indirectly* connected through one or more P1s; this represents Ts second-degree connections, which we refer to as *second-degree perceivers* (*P2s*). One can imagine even more positions further out (e.g., P3s, P4s, etc.) are theoretically possible, but we limit our scope to these three positions (T, P1, and P2) of the reputational network in the current research. As depicted in Figure 1, we refer to impressions held by the collection of P1s as the target's *direct reputation*, since information flows directly from T to P1s. We refer to impressions held by P2s as the target's *hearsay reputation*, since information flows indirectly from T to P2 through one or more P1s, via the process of gossip or hearsay.

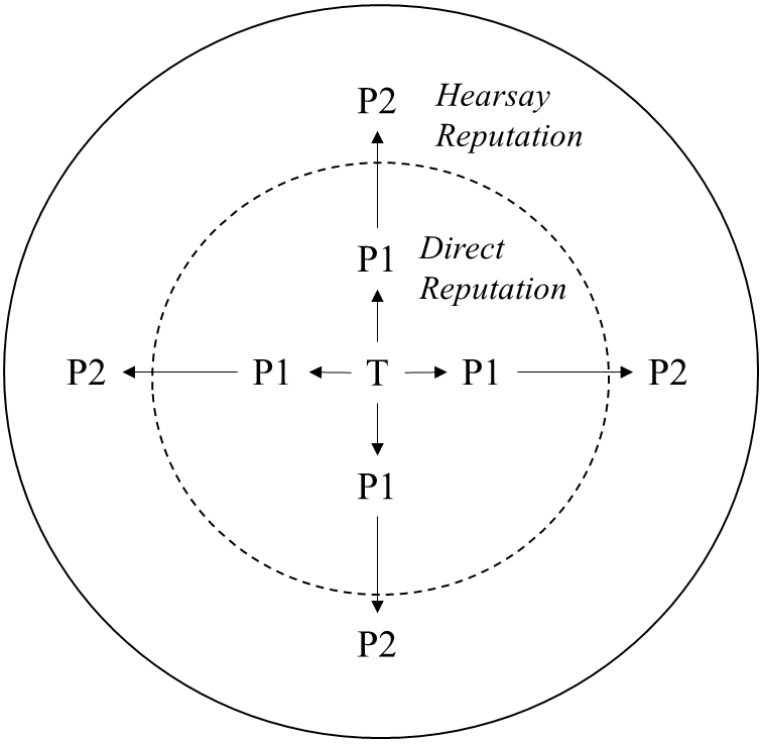


Figure 1. Graphical depiction of a generic reputational network. T = Target; P1 = first-degree Perceiver; P2 = second-degree perceiver.

The network approach to reputation thus distinguishes between 3 positions in the network, Ts, P1s, and P2s, which can be contrasted with the dyadic approach's distinction between just 2 positions, target and perceiver (also sometimes referred to as self and other, or identity and reputation; see Funder, 1995; Hogan, 1996; Kenny, 1994, 2004; McAbee & Connelly, 2016; Vazire, 2010). Dyadic approaches to interpersonal perception do not treat all perceivers equally, recognizing effects of acquaintanceship and relationship types (e.g., friend, parent, etc.) on perceivers' accuracy. In our extension, acquaintanceship and relationships are well captured by the network concepts of *tie strength* (i.e., the strength of connection between two nodes) and *multiplexity* (i.e., having ties in multiple networks; e.g., Ts reputational and familial network in the case of a parent) respectively (Kadushin, 2012). The network approach can distinguish between perceivers in terms of positionality (i.e., the number of edges on the shortest path from P to T; this is the distinction between first-degree and second-degree

perceivers), tie strength (acquaintanceship), and multiplexity (having ties in multiple networks). Positionality is the central concern of the present studies.

The concept of reputations as information stored in nodes and transmitted through edges, and the distinction between direct and indirect reputational information, is central to the network approach. We develop here a conceptual framework and empirical paradigm for studying information contained in the nodes – in other words, the impressions held by perceivers at different positions in a reputation network. By comparing the impressions at each position – impressions held by Ts (of their own self), P1s, and P2s – we can infer how information has flowed between those nodes. This follows the rich tradition of models within the dyadic approach like Funder's (1995) Realistic Accuracy Model (RAM), Vazire's (2010) Self-Other Knowledge Asymmetry (SOKA) model, and Kenny's (1994) PERSON model, comparing judgements by different perceivers to infer the underlying process of information flowing directly from T to P1. We build on this tradition by moving beyond T-P1 dyads and examining indirect information shared among perceivers[1]. By extending to P2s, the network approach necessarily focuses on indirect information, because it is the only information P2s can access. The present studies provide the first account to our knowledge of impressions based only on indirect information, what we call hearsay reputations.

Another change from the dyadic to network approach is the importance of triads (vs. dyads) in the network approach. Triads are often considered the basic building block of social networks more generally (Kadushin, 2012), and the T-P1-P2 triad is a fundamental unit of analysis for the network approach to reputations (analogously to Target-Perceiver dyads in the

[1] The PERSON model does account for indirect information flowing between P1s, what it refers to as the communication parameter, but relatively few studies have examined it (see Chaplin & Panter, 1993; Malloy, Agatstein, Yarlas, & Albright, 1997).

dyadic approach). That is, empirical paradigms within the dyadic approach tend to be designed for recruiting (e.g., through peer nominations) or creating (e.g., in round-robin designs) a set of target-perceiver dyads, with individuals sometimes serving multiple roles across overlapping conceptual dyads (as in round robin designs, where each subject is both a target and a perceiver in multiple overlapping dyads). Analogously, the paradigm we introduce in the present studies consists of extending this from dyads to triads by creating *micronetworks* of T-P1-P2 triads. We recreate the T-P1 link by having two people get to know each other, subsequently creating the P1-P2 link by having each P1 tell a naïve person (P2) about the person they met. In both studies, we use this basic approach to investigate the valence, consistency, and accuracy of information flowing across reputation networks. In Study 2, we additionally examine whether this consistency and accuracy is moderated by the context of the P1-P2 interaction, and the consequences reputational information has on social decisions made about the target.

Assessing the Flow of Reputational Information by Extending Interpersonal Perception

The primary aim of these studies is to better understand the flow of reputational information, which we do by assessing the extent to which hearsay reputations exhibit elevation differences, differential consensus, and differential accuracy. Returning to Craik's (2009) notion that reputations are networks of people that store and transmit information, we can compare the information "stored" by people at different places in the network to understand how information traverses it. The extent to which information stored at each layer is similar or different to information stored at other layers can be used to infer the valence, volume, and validity of information flowing across reputation networks in a way analogous to how we have learned about the valence (e.g., self-enhancement; Kwan, John, Kenny, Bond, & Robins, 2004), consistency (e.g., interjudge consensus; Connelly & Ones, 2010; Kenny, 1994; Funder, Kolar, &

Blackman, 1995), and accuracy (Funder, 1995; Kenny, 1994; Vazire, 2010) of direct reputations by comparing P1s' impressions to each other and to Ts' self-reports.

Analogous to the dyadic approach, there are several ways to compare Ts', P1s', and P2s' perceptions, each of which speaks to different qualities of the judgments. One way is to compare the means of perceptions in each layer. This speaks to any possible tendency for some layers of the network (self-perception, direct reputations, or hearsay reputations) to be systematically different (e.g., more positive or negative) than others. Following Cronbach's (1955) use of elevation to refer to average ratings made by judges, we refer to differences between judges (selves, P1s, P2s) as *relative elevation*. A second way to assess the transmission of information in reputation networks is to evaluate the extent to which the relative ordering of ratings is similar across nodes in different layers (i.e., estimate the correlations between Ts, P1s, and P2s), which following Cronbach we call *differential consensus and accuracy*. We now discuss each in greater detail.

Relative elevation: changes in valence of information flow. How does the valence of reputational information – the relative positivity or negativity – change as it travels along the edges from targets, to P1s, P2s, and beyond? In dyadic studies, there is a frequently observed tendency for individuals to self-enhance for ability-related and more evaluative personality attributes (John & Robins, 1993, 1994; Kwan et al., 2004; Leising, Erbs, & Fritz, 2010; Robins & John, 1997; but see Leising, Locke, Kurzius, & Zimmermann, 2016 for exceptions). Put differently, direct reputations are less positive than self-perceptions, suggesting that information becomes less positive as it travels along the edge from target to P1.

We did not have strong a priori hypotheses about hearsay reputations, but we considered several possibilities. First, it is possible that the positivity of an impression is influenced by the

distance between the target and perceiver, with reputational information becoming less positive as it travels from P1s to P2s; we will refer to this as the *distance hypothesis* of relative elevation. This would be consistent with various work demonstrating that we tend to like and think positively of those that we know well and are close too (e.g., Leising et al., 2010; Murray, Holmes, & Griffin, 1996; Vazire, 2010). If this is the case, we would find positive self-P1 relative elevation (that is, self more positively valenced than P1), positive P1-P2 relative elevation, and positive self-P2 relative elevation.

Second, consistent with conceptualizations of self-enhancement emphasizing the unique motivation to maintain positive self-perceptions (e.g., Dunning, 1999), it is possible that positivity is not a function of distance from the target but instead a function of self vs. (most) others, and that reputational information remains similarly positive as it travels from P1s to P2s; we will refer to this as the *self-enhancement hypothesis*. This would result in positive self-P1 and positive self-P2 relative elevation.

Third, it is possible that P1s disproportionately share positive information about targets with P2s, stemming from the social cost of sharing negative gossip (Craik, 2009; Farley, 2011), leading to hearsay reputations being more positive than direct reputations; we'll refer to this as the *positively biased hearsay hypothesis* of relative elevation. This would result in negative P1-P2 relative elevation (P1 less positively valanced than P2).

Differential consensus and accuracy: strength of information flow. *Differential consensus and accuracy* are estimated by correlating vectors of ratings made by different judges and each assesses the extent to which the relative ordering of ratings made by one set of judges (e.g., P1s) corresponds to the relative ordering of ratings made by another set of judges (e.g., self-ratings by Ts). When assessed among only perceivers (Ps), the correlation(s) are referred to

as consensus (Funder, 1995; Kenny, 1994). When one of the ratings is a self-rating by the target (T), the correlations are often referred to as self-other agreement. Sometimes this is also referred to as accuracy when there is a defensible argument that the self can be treated as an accuracy criterion (Kenny, 1994). Research on differential consensus and accuracy have provided valuable insights into interpersonal perception, generally finding some degree of consensus and accuracy across a broad range of contexts (Back et al., 2010; Connelly & Ones, 2010; Funder, 1995; Gosling et al., 2002; Kenny, 1994).

Differential consensus and accuracy have been central to Funder's (1995) accuracy paradigm of personality judgements and Kenny's (1994) interpersonal perception paradigm, and so each have been examined extensively across a range of contexts. Meta-analytic evidence suggests that both inter-judge (P1-P1) consensus and self-other (T-P1) agreement are moderate, though inter-judge consensus tends to be greater than self-other agreement (Connelly & Ones, 2010). Importantly, both consensus and self-other agreement vary considerably depending upon which personality trait is being perceived, the source of information (e.g., photograph, face-to-face interaction, etc.), and the relationship between the judge and target. This variability is expected under Funder's (1995) Realistic Accuracy Model (RAM), wherein accurate personality judgment is thought to vary according to features of the judge, target, trait, and information.

Differential consensus and accuracy in hearsay reputations. A network approach includes different kinds of *hearsay consensus*, including P2-P2 consensus and P1-P2 consensus. In the present study we only examined P1-P2 consensus, and so we will use the term "hearsay consensus" interchangeably with P1-P2 consensus. Hearsay consensus speaks to the strength of information flowing from P1 to P2. Hearsay accuracy refers to the extent to which hearsay reputations are accurate, and simultaneously speaks to the validity of the information flowing

from P1 to P2 and to the strength of information flowing from the Target to P2 (indirectly through a P1). There is no single, perfect criterion to assess the accuracy of personality judgments (see Funder, 1995; Kenny, 1994; Robins & John, 1997); we use targets' self-reports as the accuracy criterion in this study, as is commonly done (Back et al., 2010, German Sample; Biesanz, 2010; Biesanz & Human, 2010; Graham & Gosling, 2012; Harari, Graham, & Gosling, 2015; Human & Biesanz, 2011, 2012), because we believe it's reasonable to privilege self-perception of Big Six personality domains over impressions made by two new acquaintances. We thus refer to self-P2 agreement as *hearsay accuracy* and self-P1 agreement as *direct accuracy*, since each measures the accuracy of hearsay and direct reputations respectively.

Based on the previous work outlined above, we expect to find some degree of hearsay consensus and accuracy. We anticipate differences between hearsay consensus and accuracy such that the former is greater than the latter, mirroring findings with respect to self-P1 and P1-P1 agreement (see Connelly & Ones, 2010). One reason P1-P1 agreement exceeds self-P1 agreement is that P1s share many of the motivational and informational factors that might bias perceptions with each other, but not with the self (Vazire, 2010). Following this logic, hearsay consensus might be lower than P1-P1 agreement typically is; P1s and P2s may have different motivations, and certainly have access to different kinds of information (direct vs. indirect), each of which could dampen hearsay consensus.

However, agreement between connected P1s and P2s may be even greater than benchmarks of P1-P1 agreement. Kenny's (2004) PERSON model conceptualizes an interpersonal judgment as a combination of six different components. PERSON contains three components of stable information in perceivers' judgments: stereotype, personality, and opinion. Stereotype and personality, as defined in PERSON, are two sources of stable information that

would be shared across all perceivers (based on appearance and behavior, respectively) – they are the main mechanisms underlying P1-P1 consensus. Opinion is one perceiver's information that is stable but idiosyncratic: it is *not* shared with other perceivers, and it can be thought of as that perceiver's unique take on the target, which does not promote P1-P1 consensus. Kenny's (2004) work with the PERSON model suggests that opinion comes to dominate any single perceiver's impression, and perceivers are unaware of what part of their impression belongs to each component. If P1s share all three aspects with P2s – which seems plausible given that they cannot distinguish between them - then agreement among connected P1s and P2s may be even greater than what is typically found among P1s, since they share the opinion component in addition to the components generally shared among perceivers (personality and stereotypes). We might thus expect hearsay consensus to be greater than both hearsay accuracy and previous estimates of P1-P1 consensus.

We also expect personality domains to differ in the extent of hearsay consensus and accuracy, mirroring previous work on direct reputations (i.e., self-P1 agreement and P1-P1 agreement). Extraversion tends to be the domain with the greatest extent of consensus and accuracy in face-to-face interactions, which characterizes the T-P1 interaction in the present studies (see Connelly & Ones, 2010; see also Kenny, 1994). However, we approached as an open question how well this generalizes to the medium of hearsay. For example, it is possible that P1s have more information about targets' extraversion – which is implied by previous work demonstrating greater direct consensus and accuracy for extraversion – but for whatever reason focus less on extraversion when describing the target. Give this uncertainty, we did not make specific predictions about which personality domains would show the greatest extent of hearsay consensus and accuracy.

What Moderates the Flow of Reputational Information?

A central goal of both studies is to estimate hearsay consensus and hearsay accuracy for the first time, with the broader aim of better understanding how reputational information flows across networks of people. We expect the flow of reputational information, and therefore the extent of hearsay consensus and accuracy, to vary across the Big Six domains. Differences across the Big Six could occur because P1s don't have information to share about a domain (e.g., if they have little or no information about a target's neuroticism) or because they don't think it is relevant information to share (e.g., they perceived the target as anxious but simply don't think it's worth bringing up). These follow what the Linguist Paul Grice (1975) referred to as the conversational maxims of quality ("do not say that for which you lack adequate evidence" p.46) and relevance or relation ("be relevant" p.46) respectively. We attempt to affect relevancy specifically in Study 2, but this distinction remains to be studied more closely in the future.

Building on work examining good judges of personality and good (easily judged) targets (Funder 1995; Human & Biesanz, 2013), we expected that individual differences among targets, P1s, and P2s could impact hearsay consensus and accuracy. We specifically examined extraversion (among targets, P1s, and P2s) and empathy (among P2s only) as moderators of hearsay consensus and accuracy in both studies. Targets' extraversion has been found to promote accurate interpersonal perception, presumably via extraverts' heightened expressivity resulting in a greater number of cues available to perceivers (Funder, 1995; Human & Biesanz, 2013). Analogously, P1s' extraversion may influence how much they say about the target, ultimately affecting the amount of information available to P2, impacting hearsay consensus and accuracy in the process. We might expect P2s' extraversion to operate similarly, increasing the information at their disposal by engaging more with P1. However, if extraversion's impact on

interpersonal perception is solely due to making more cues available, we may expect P2s' extraversion to have little impact (since they do not have any information about targets). On the more exploratory side, we examined P2s' empathy (the four domains within the Interpersonal Reactivity Index; Davis, 1983) as a moderator based on the intuition that more empathic P2s may have an easier time understanding P1s perception of the target.

A variety of factors may influence hearsay consensus and accuracy, including tie strength (i.e., acquaintanceship) between individuals (T-P1, and P1-P2), multiplex relationships (e.g., romantic ties, familial ties, etc.), goals to convey certain impressions (i.e., a reputational management goal, analogous to impression management for the self), and the context of the target-P1 interaction. A systematic examination of all these potential factors was well beyond the scope of a single paper. However, as a beginning step, we systematically studied the context of the hearsay exchange between P1 and P2 in Study 2.

Context and hearsay reputations. We propose that one important part of the context of the hearsay exchange – the conversational goal that was provided to the participants – could affect hearsay consensus and accuracy. We based this on two principles. The first is Grice's (1975) maxim of relation or relevance, which holds that conversational norms require individuals to provide information that is relevant to the topic at hand. The second is the idea that personality domains are not equally relevant across contexts (Cottrell et al., 2007; see also Buss, 2011; Srivastava, 2010; Wood, 2015). Taken together, this suggests that individuals may provide more information about personality domains that are relevant to the context in which the hearsay discussion takes place. For example, hearsay in a work context (e.g., asking a reference about an applicant) may focus disproportionately on characteristics relevant to the job (e.g., the domain of conscientiousness), whereas telling a friend about a new romantic interest may focus on

information more relevant to romantic relationships (e.g., the domain of agreeableness). The context of the hearsay exchange may thus impact the volume and quality of information shared about characteristics perceived to be relevant to that context, ultimately impacting the extent of hearsay consensus and accuracy for context-relevant personality domains.

The specific contexts we examined in Study 2 were conversational goals: Participants were instructed that the goal was to evaluate the target's suitability for status- or affiliation-related roles (leadership and friendship respectively). Status and affiliation, or getting ahead and getting along, are thought to be two critical and ubiquitous demands of social living (Hogan, 1996). The relation between personality traits and these contexts have been the focus of prior research and theorizing. Within the Big Five, extraversion tends to be the most consistently associated with status or leadership; extraversion tends to predict both status attainment and leadership ability (Anderson, John, Keltner, & Kring, 2001; Judge, Bono, Ilies, & Gerhardt, 2002; Lawless Desjardins, Srivastava, Küfner, Back, 2015). Affiliation, on the other hand, is theorized to be most strongly connected to agreeableness, both in terms of intimate (affiliative) relationships and getting along or cooperating with others more generally (Buss, 1996, 2011; Denissen & Penke, 2008; DeYoung, 2015). Supporting this notion, people tend to like others higher in agreeableness (Wortman & Wood, 2011), and report valuing agreeableness in friendships and other intimate (affiliative) relationships (Cottrell et al., 2007; Wood, 2015). Investigating reputations formed from hearsay in different contexts can simultaneously inform of us of the relative importance of personality domains in different contexts and how relevance affects the flow of reputational information.

The Present Studies

The present studies were designed to provide a first look at hearsay reputations. We present a laboratory paradigm and data-analytic framework that extends methodological and analytic approaches designed for interpersonal perception (e.g., Kenny, 1994; Olsen & Kenny, 2006) to study hearsay reputations among previously unacquainted individuals. In the procedure two people get to know one another. Then one or both of them tell new participants about the person they met, providing an experimental instantiation of a basic reputational network (as depicted in Figure 1). In Study 1, we provide initial estimates of relative elevation, hearsay consensus, and hearsay accuracy, and we examine individual differences that may affect hearsay consensus and accuracy. Study 2 was designed to replicate and extend Study 1 and had three basic aims. First, in Study 2 we examined the extent to which the context of the hearsay exchange affected hearsay reputations, specifically focusing on hearsay consensus and accuracy. Second, we examined social consequences of hearsay reputations, specifically how social decisions made about the target were influenced by hearsay reputations of Big Six personality characteristics. Third, we obtained more precise estimates of the analyses in Study 1 by replicating Study 1's analyses on a combined (Study 1 and Study 2) dataset.

Study 1

Study 1 was our first use of the micronetwork paradigm to estimate relative elevation, differential consensus and accuracy, and individual-difference moderators.

Method

Pre-registration. The measures and materials, procedure, sample size, exclusion criteria, and analysis plan were pre-registered on the Open Science Framework (OSF; pre-registration: https://www.osf.io/hqtfb; full project site: https://www.osf.io/65wfu/). The full procedure (described below) was carried out as described in the pre-registration. Our pre-registered

sampling plan was to collect data until we had 100 targets (between 50 and 100 groups of 3 and 4 participants), and to continue collecting data beyond this point if it was minimally costly (e.g., continuing data collection until the end of the academic term). Our pre-registration also included a blinded screen (MacCoun & Perlmutter, 2015) whereby a member of the research team (the second author) screened de-identified and unlinked datasets for invalid responding, such as long runs of the same response (e.g., many 3's in a row) or many skipped questions. In our initial pre-registration, we accidentally omitted an important exclusion criterion: that groups would be excluded if anyone knew one of the people they met or heard about. We added an addendum to our pre-registration shortly after data collection began but before we had looked at the data; the addendum is also on OSF (https://www.osf.io/eckj7).

Analyses were pre-registered as primary or secondary, though they may be more properly called higher and lower priorities. That is, our main distinction between primary and secondary analyses is that we viewed the latter as less central and we thus committed to reporting all primary analyses but only select secondary analyses. Primary analyses included hearsay consensus and accuracy for the Big Six. Secondary analyses included meta-accuracy for the Big Six, hearsay consensus and accuracy for traits outside of the Big Six that we measured (e.g., self-esteem, physically attractive, etc.), and individual difference moderators (each triad member's extraversion and P2s' Empathy). For the sake of space, we present the primary analyses and individual difference moderators; facets, traits outside of the Big Six, and meta-perceptions will not be included in this manuscript. We additionally report several (non-pre-registered) analyses of additional interest; because the majority of analyses were pre-registered, we indicate which analyses were *not* pre-registered. As noted in our pre-registration for Study 1, our focus is on effect size and precision, but we also conducted two-tailed significance tests using alpha = .05 as

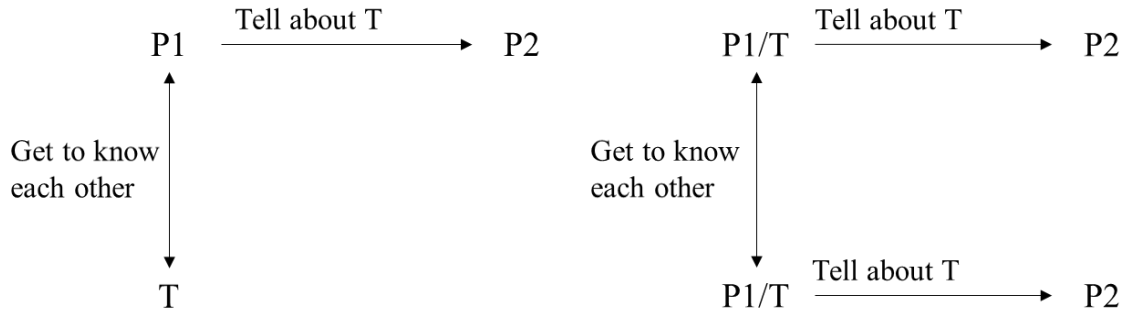
our cutoff; both criteria were pre-registered. We present 95% confidence intervals around each parameter we estimate, and to aid interpretation, we compare the effect sizes we observe in our study to benchmarks from previous literature (specifically, meta-analytic estimates taken from Connelly & Ones, 2010). We present exact *p* values and use significance tests to help rule out that the correlations arise from chance guessing alone (e.g., zero consensus or accuracy). The study was approved by the University of Oregon's institutional review board.

Participants. Participants were recruited for the study from the University of Oregon Human Subjects Pool and were unaware of the topic of the study upon signing up. Participants signed up independently and participated in groups of 3 or 4. Our initial sample consisted of 288 participants. Groups of participants were excluded from analyses if any participant indicated that they knew one of the people they interacted with; this resulted in eliminating 7 groups. An additional group was unable to finish the study due to a mandatory evacuation in the building. Finally, 3 additional participants were excluded after they reported to the researcher(s) that they did not follow the questionnaire instructions. This resulted in a final sample of 260 participants in 72 groups. The final sample included 115 participants in the target role, exceeding our pre-registered minimum goal of 100 targets.

Participants were aged 18 to 30, with an average age of 19.6 (SD = 1.8). The majority of participants identified as Female (58% of those that reported; 15% did not report gender). The majority of participants identified as White/Caucasian (60%), followed by Asian (18%), Hispanic/Latino (9%), other (4%), Black/African American (4%), Native American (< 1%), and 15% did not report race/ethnicity; participants were able to select more than one option.

Procedure. As noted above, the procedure was pre-registered on OSF, and the full procedure, script, and instructions can be found at the pre-registration link

(https://www.osf.io/hqtfb). Participants were scheduled in groups of 3 or 4. Participants were instructed to arrive and wait in separate pre-determined areas while awaiting the experimenter to ensure that they had little to no contact prior to the start of the experiment. Before the start of the experiment, participants were randomly assigned to the different procedural 'roles', which corresponded to one or more of the network positions outlined above (T, P1, P2). Figures 2a and 2b show how the triads (Figure 2a) and tetrads (Figures 2b) of participants map onto the theoretical roles of interest. As shown in Figure 2a, each participant in a three-person group served one role, completing a single reputational triad[2], with one participant serving the target role, one serving the P1 role, and one serving the P2 role. As shown in Figure 2b, four-person groups were more complicated; two participants were simultaneously targets and P1s and two participants were P2s (see Figure 2b). Four-person groups thus contained two overlapping reputational triads. Experimenters retrieved participants, one at a time, and directed them to an empty lab room where they were instructed to read the informed consent document and await the start of the experiment. From this point, the experiment proceeded in three phases: 1) dyadic interaction, 2) hearsay, and 3) ratings.



(a) Three-person group (b) Four-person group

Figures 2a-b. Graphical depiction of study design. Participant roles appear in parentheses below the name.

[2] The target and P1 did rate each other, making them simultaneously a target and perceiver of each other. However, each played a single role with respect to the T-P1-P2 reputational triad.

Phase 1: Dyadic interaction. Two participants were brought into a lab room and sat at opposite ends of a rectangular table that had a stack of cards at the center of it. The experimenter then asked if the participants knew one another; the experiment continued regardless of their response, but "yes" answers were recorded, and those sessions were later excluded from analyses. Then, the experimenter instructed them that they were to play a get-to-know-you game, in which participants took turns asking questions printed on the cards to which both were expected to respond. Questions were geared towards revealing broadly relevant personality information. We included 12 questions that were each targeting a Big Six domain (2 questions per domain). For example, "What do you do to manage schoolwork and deadlines?" was intended to target conscientiousness. We included 6 additional questions eliciting small talk (one from the control condition of Aron et al., 1997). Participants were instructed to go through the questions in order and to get through as many of the questions as they could in the 15 minutes they were allotted. These conversations were recorded, and participants were aware of these recordings. Note that the two participants in this initial interaction had procedurally identical roles; assignment as a T, a P1, or (in a 4-person group) a dual T/P1 was not relevant to what they did.

Meanwhile, the remaining participant(s) were seated in front of their own computer (alone) and completed a questionnaire consisting of the Interpersonal Reactivity Index (Davis, 1983) and filler questions for 15 minutes.

Phase 2: Hearsay. In the second phase, one or both of the participants from the phase-1 interaction had a conversation with a new participant about the person they had just met. In sessions run with 3-person groups, one person from the phase-1 interaction was assigned to the role of P1 and met with a new participant serving as a P2. In 4-person groups, each participant

from the phase-1 interaction now served as a P1 describing their previous interaction partner to a new P2. P1s were instructed to tell the new person (P2s) about the person whom they had just met. They were instructed to, "…give the most informative, objective, and accurate account of what the person you [they] met was like," and P2s were similarly instructed to, "try to get the most accurate impression you [they] can from what [P1] tells you [them]." Full instructions can be found on OSF (https://www.osf.io/hqtfb). These conversations were also recorded; participants were aware of the recordings.

Phase 3: Ratings. After approximately 10 minutes had passed, participants were separated and instructed to complete a series of questionnaires on the computer. After finishing, participants were thanked and debriefed.

Measures. Questionnaires included self-, P1-, and P2-reported personality questionnaires as relevant to each participant's roles (described below), followed by a brief set of demographic questionnaires. Participants in the P2 role additionally completed open-ended descriptions of the person they heard about; these are not analyzed for this manuscript.

Big Six. We administered self-, P1-, and P2-reports of the Big Six personality domains. We used the full (60-item) Big Five Inventory 2 (BFI-2; Soto & John, 2017) to measure the Big Five domains, which are agreeableness, conscientiousness, neuroticism, extraversion, and openness. We used eight items from the 48 Questionnaire Big Six (48QB6; Thalmayer, Saucier, & Eigenhuis, 2011) to measure honesty-propriety. All ratings were completed on a scale ranging from 1 (Disagree Strongly) to 5 (Agree Strongly) with a neutral point of 3 (Neutral; no opinion). The only difference between self-, P1-, and P2-reports were the instructions, which were appropriately re-worded to reflect the target they were rating.

Coefficient alphas for each of the domain scales of the Big Six are reported in Supplemental Table S1 and summarized here. Among full scales, honesty-propriety consistently had the lowest alpha value (.61, .66, and .75 for self, P1-, and P2-reports respectively); this may be because it had the fewest items (8 for honesty-propriety vs. 12 for each of the Big Five). All other alpha coefficients were above .70.

Empathy. Participants in the P2 role completed a questionnaire while they waited for the second phase. The questionnaire mostly consisted of filler questions about their social connectivity (e.g., number of contacts in their phone, number of Facebook friends, number of text messages, etc.). The questionnaire also included a measure of empathy, the Interpersonal Reactivity Index (IRI; Davis, 1983), which measures four aspects of empathy (perspective taking, personal distress, fantasy, and empathic concern) with 28 short statements (e.g., "I believe that there are two sides to every question and try to look at them both" from Perspective Taking) that are rated on a 0 (Does not describe me well) to 4 (Describes me very well) scale. All four subscales demonstrated adequate internal consistency (α's from .73 to .82).

Analytic approach. In this design, individuals are nested in groups, creating 3- or 4-person reputational micro-networks designed to emulate a reputational network (Figure 1 depicts a theoretical reputational network; Figures 2a and 2b maps the reputational network onto the 3- and 4-person micro-networks). Our data analytic approach draws on concepts and principals from dyadic data analysis (e.g., Actor Partner Interdependence Models; APIM; Olsen & Kenny, 2006) and applies them to 3- and 4-person micro-networks. Two concepts were central to our approach. First is the concept of dependencies. APIM deals with dyadic dependencies by treating effects as nested within dyads; we similarly dealt with dependencies in the 3- or 4-person micro-networks by treating effects as nested in triads and tetrads. Second is the concept of pooling

interchangeable estimates. In APIM, this means pooling means, variances, and paths (regression paths representing actor and partner effects) across dyad members. Our design is more complicated; put succinctly, we pooled estimates whenever people occupied the same theoretical role (T, P1, or P2) *and* were procedurally exchangeable (i.e., did the same thing in the experiment). Supplemental Table S2 contains an example variance-covariance matrix with equality constraints noted in its cells.

Analyses were run in the *lavaan* package (Version 0.6-3; Rosseel, 2012) and a wrapper R package created by the first author (see: https://github.com/coryc3133/ReputationModelR) in R (version 3.6.1; R Core Team, 2019) and additionally verified in Mplus (version 7.31; Muthén & Muthén, 1998-2012). Relative elevation and differential consensus/accuracy results were obtained by modelling the full variance-covariance and mean structure of the ratings for each of the Big Six with the appropriate equality constraints to pool exchangeable parameters (outlined above). Individual difference moderators and distinctive hearsay accuracy were assessed in path analytic regression models (with analogous equality constraints). In all analyses, we used full information maximum likelihood to treat missing data; this was necessary given the planned missingness in our design (i.e., including both three- and four-person groups). The data and a script that reproduces the analyses can be found at our OSF project page (https://www.osf.io/65wfu).

Results

Relative elevation. Relative elevation was analyzed by estimating differences in means for each of the Big Six domain scales separately. For each comparison, Table 1 contains an estimated raw difference as an unstandardized effect size (in the original 1 to 5 Likert-type units), a 95% CI around the raw difference, a standardized effect size (Cohen's *d*), and a *p* value.

The pooled means and standard deviations for each rating (perceptions of the target by self, P1, and P2) can be found in Supplemental Table S3.

Table 1

Relative Elevation Differences, Study 1

		D	95% CI	*d*	*p*
P1-P2	A	0.30	[0.19, 0.41]	0.58	<.001
	C	0.14	[0.02, 0.27]	0.25	.020
	HP	0.39	[0.28, 0.50]	0.76	<.001
	N	-0.22	[-0.31, -0.12]	-0.40	<.001
	E	-0.04	[-0.16, 0.08]	-0.05	.537
	O	0.15	[0.05, 0.25]	0.32	.003
Self-P2	A	0.18	[0.06, 0.31]	0.33	.005
	C	-0.06	[-0.20, 0.08]	-0.09	.398
	HP	0.29	[0.16, 0.41]	0.51	<.001
	N	0.16	[0.00, 0.32]	0.22	.052
	E	0.20	[0.04, 0.35]	0.27	.014
	O	0.50	[0.38, 0.61]	0.92	<.001
Self-P1	A	-0.12	[-0.22, -0.02]	-0.22	.021
	C	-0.20	[-0.31, -0.09]	-0.34	<.001
	HP	-0.10	[-0.21, 0.00]	-0.19	.052
	N	0.38	[0.23, 0.52]	0.54	<.001
	E	0.23	[0.11, 0.36]	0.32	<.001
	O	0.34	[0.25, 0.44]	0.65	<.001

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. P1-P2 = P1(T) – P2(T); Self-P2 = T(T) – P2(T); Self-P1 = T(T) – P1(T). D denotes raw differences; *d* denotes Cohen's *d* values. 95% CIs correspond to raw differences. Values are obtained from a model pooling means and variances for exchangeable triad members.

P1-P2 relative elevation. Did perceivers who only heard about the targets form different impressions than perceivers who directly interacted with them? We compared the relative elevation of direct reputations and hearsay reputations separately for each of the Big Six. As seen in Table 1, P2s consistently rated the target less positively than P1s did for each of the Big Six except extraversion (which showed no difference). Taking agreeableness as an example, the difference score of 0.30 indicates that P1s rated the target more Agreeable than P2s did by about a third of a Likert-scale point, and anything from one- to two-fifth of a scale point is within the 95% confidence band. The Cohen's *d* of 0.58 means that a P1 and P2 rating the same target's agreeableness are expected to differ by almost two-thirds of a standard deviation. Across traits, differences ranged from a high of $d = 0.76$ for honesty-propriety to a low of $d = 0.25$ for

conscientiousness, which cover Cohen's range from large to small effects. Moreover, the CIs include positive but negligible differences for conscientiousness and openness (0.02 and .05 in raw scale units, respectively). In summary, hearsay reputations thus tended to be less positive than direct reputations, but there was some heterogeneity in the size of these differences across Big Six domains.

Self-P2 relative elevation. Next, we turn to differences between hearsay reputations and the self. As seen in Table 1, P2s tended to rate the target more negatively than the targets rated themselves, with the exception of neuroticism and conscientiousness, neither of which was significant. Outside of neuroticism and conscientiousness, discrepancies ranged from $d = 0.92$ for openness to $d = 0.27$ for extraversion. For openness, this means that self- and P2-reports of the same target are expected to differ by almost a full standard deviation. For extraversion, self- and P2-ratings are much closer together, differing just a quarter as much as the average between-target difference. Moreover, the lower bounds of the CIs indicate that our evidence is also consistent with a negligible difference for extraversion and agreeableness (.04 and .06 respectively; see Table 1). Hearsay reputations thus tended to be less positive than self-reports, but that pattern was less consistent across Big Six domains than the P1-P2 relative elevation results.

Self-P1 relative elevation. We also compared the relative elevation of self-perceptions and direct reputations – the classic approach to self-enhancement. As seen in Table 1, Self-P1 relative elevation was somewhat inconsistent across domains. Extraversion and openness followed the classic self-enhancement pattern, where P1s rated targets more negatively than targets rated themselves by roughly one- and two-thirds of an SD, respectively. However, the opposite pattern held for agreeableness, conscientiousness, honesty-propriety, and neuroticism;

these self-effacement effects ranged from small and consistent with a negligible difference (e.g., honesty-propriety and agreeableness), to moderate (e.g., conscientiousness), to large and consistent with moderate-to-large differences (e.g., neuroticism).

Differential consensus and accuracy. Differential consensus and accuracy concern the extent to which the relative ordering of individuals is correlated across judges. We start by detailing our findings for hearsay consensus, hearsay accuracy, and direct accuracy. In keeping with our pre-registration, we use significance tests to help rule out chance guessing, but focus more on interpreting estimated effect sizes. To aid with the latter, we plot our estimated effect sizes alongside relevant meta-analytic benchmarks to aid interpretation.

Hearsay consensus. We assessed the extent to which there is differential hearsay consensus – that is, agreement between direct and hearsay reputations – by correlating P1s' ratings of the target with P2s' ratings of the target. We did so separately for each of the Big Six. This tells us the extent to which the relative ordering of people on a given dimension is preserved across different layers of a reputational network (i.e., from direct reputations to hearsay reputations). Figure 3 depicts hearsay consensus for each of the Big Six. As an interpretive benchmark, we have also plotted a meta-analytic estimate of interjudge consensus (P1-P1 consensus). The benchmark comes from a meta-analysis by Connelly and Ones (2010); it is the average interjudge correlation for studies in which ratings were completed after previously unacquainted participants interacted naturistically in a laboratory setting. This helps contextualize estimates of hearsay consensus by comparing them to more typical assessments of consensus within direct reputations.

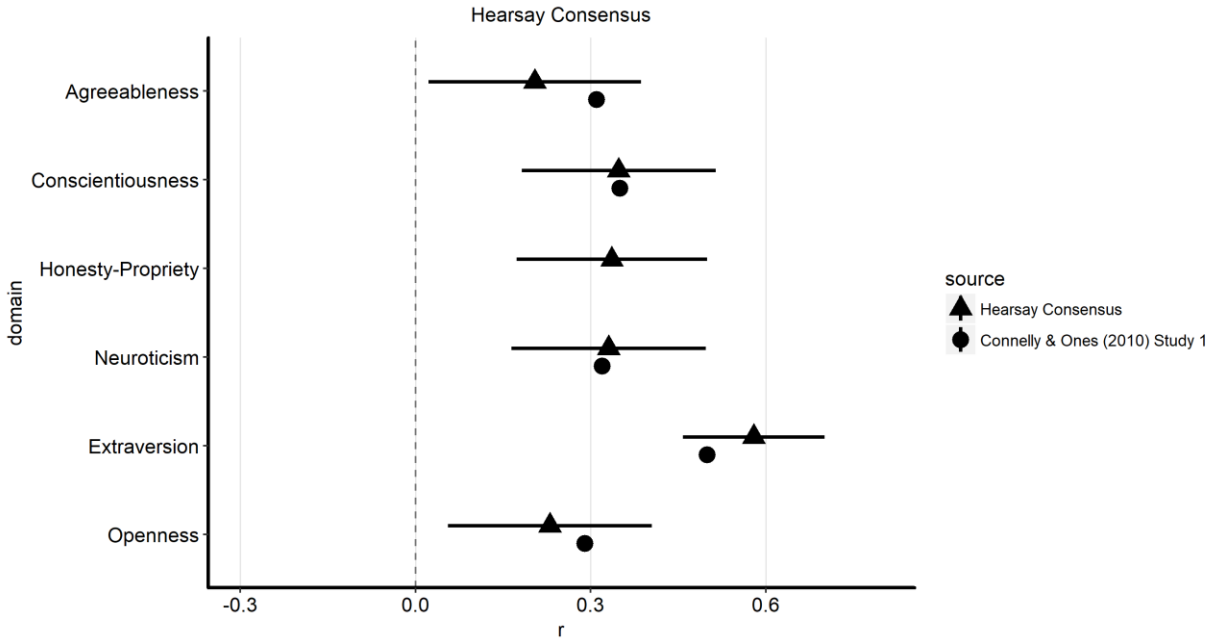


Figure 3. Hearsay consensus for the Big Six in Study 1. Triangles represent point estimates for correlations and the lines through them represent the 95% CI's around those estimates. Circles represent meta-analytic estimates of interjudge consensus following naturalistic laboratory interactions, obtained from Study 1 of Connelly & Ones's (2010) meta-analysis on consensus and self-other agreement.

Figure 3 reveals that there is a moderate amount hearsay consensus for all of the Big Six, though estimates vary across trait domains. Hearsay consensus was lowest in magnitude, though still distinguishable from chance guessing, for agreeableness (r = .20, 95%CI [.01, .39], p = .036). It was moderately high for conscientiousness (r = .34, 95% CI [.18, .51], p < .001), honesty-propriety (r = .32, 95%CI [.16, .49], p < .001), and neuroticism (r = .34, 95%CI [.18, .50], p < .001), quite high for extraversion (r = .57, 95%CI [.45, .70], p < .001), and fairly low (though distinguishable from chance guessing) for openness (r = .23, 95%CI [.05, .41], p = .011). Moreover, each estimate was quite similar to meta-analytic estimates of interjudge consensus (between P1s). This suggests that hearsay consensus (i.e., P1-P2 agreement) was similar in magnitude to direct consensus (i.e., P1-P1 agreement) seen in previous work. It thus appeared

that P1s were able to communicate a consistent impression of the target to P2s – indeed, an impression roughly as consistent as if both perceivers had met the target themselves.

Hearsay accuracy. We assessed hearsay accuracy by correlating hearsay reputations (ratings by P2s) with the targets' self-reports, treating the self-reports as an accuracy criterion. Figure 4 contains hearsay accuracy for each of the Big Six plotted alongside a meta-analytic estimate of self-other agreement among previously unacquainted participants, taken from the same meta-analysis as above (Connelly & Ones, 2010). This provides context for hearsay accuracy by allowing comparison to meta-analytic estimates of direct accuracy.

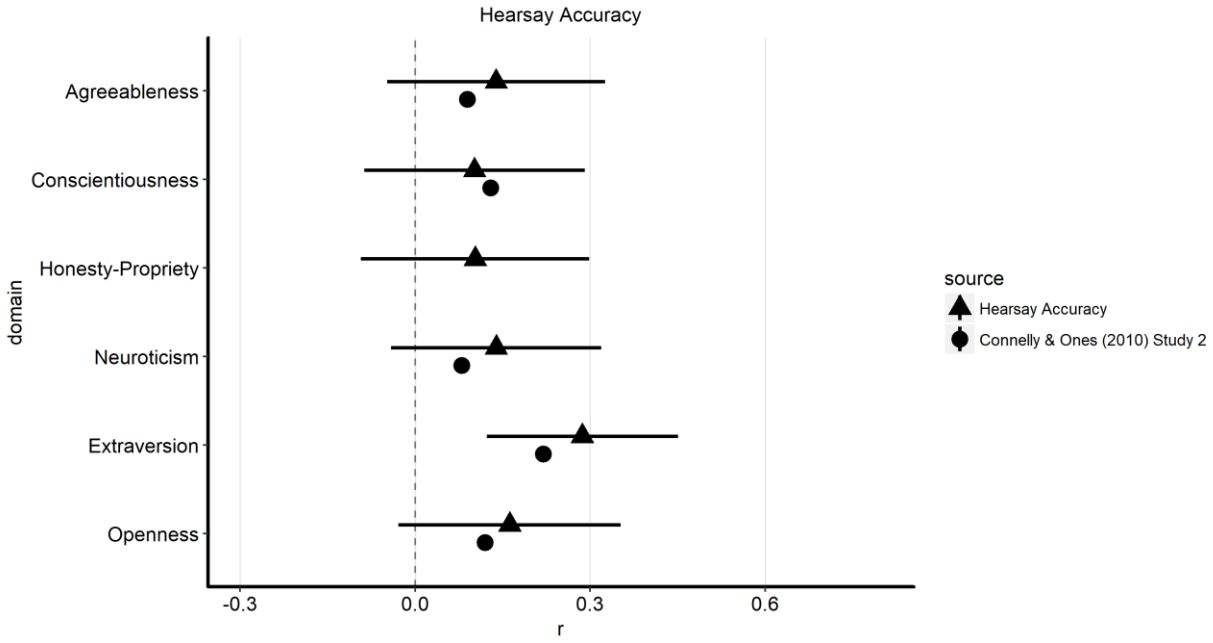


Figure 4. Hearsay accuracy for the Big Six in Study 1. Triangles represent point estimates for correlations and the lines through them represent the 95% CI's around those estimates. Circles represent meta-analytic estimates of self-other agreement among previously unacquainted participants, obtained from Study 2 of Connelly & Ones's (2010) meta-analysis on consensus and self-other agreement

As depicted in Figure 4, hearsay accuracy was low and indistinguishable from chance guessing for agreeableness (r = .17, 95% CI [-.02, .36], p = .075), conscientiousness (r = .12, 95% CI [-.07, .30], p = .229), honesty-propriety (r = .09, 95% CI [-.11, .30], p = .363), and

neuroticism (r = .14, 95% CI [-.04, .32], p = .120). Hearsay accuracy was moderately high and distinguishable from chance guessing for extraversion (r = .28, 95% CI [.12, .45], p = .002), and it was low and indistinguishable from chance guessing for openness (r = .16, 95% CI [-.02, .35], p = .085). Each estimate was similar to the corresponding meta-analytic estimates of self-other agreement, suggesting that, at least descriptively, P2s reach similar levels of accuracy as P1s when using self-reports as the accuracy criterion. Taken together, it would appear that hearsay accuracy is descriptively similar to meta-analytic estimates of direct accuracy (i.e., self-other agreement with direct perceivers), though in Study 1 we could only rule out chance guessing for extraversion.

Direct accuracy. We also assessed direct accuracy by correlating P1s' ratings with the targets' self-reports, thus treating self-reports as the accuracy criterion in these analyses as well. Direct accuracy was similar to meta-analytic estimates from Connelly and Ones (2010), though it was higher on average, with *r's* from .42 for Extraversion (p < .001) to .15 for Honesty-Propriety (p = .074; see Supplemental Figure S1). Thus, P1s achieved a similar degree of accuracy to what would be expected from prior work.

Distinctive hearsay accuracy. In an additional analysis that was not preregistered, we were interested in whether P2 impressions contained distinctively accurate information that was not reflected in P1 ratings – in other words, whether P2s knew something accurate about the target that P1s did not. This would be possible, for example, if during the hearsay interaction P1s described trait-relevant behavior that they did not interpret correctly but P2s did. This analysis was inspired by similar questions in meta-accuracy research (Carlson, Furr, & Vazire, 2010). We assessed distinctive hearsay accuracy by regressing targets' self-reports on both P1s and P2s ratings; this analysis was performed for each personality domain separately. Table 2 contains the

results from these models. As seen in Table 2, we did not find evidence of distinctive hearsay accuracy, suggesting that the limited amount of hearsay accuracy found in this study was achieved primarily through P2s adopting P1s' view of the targets.

Table 2

Distinctive Hearsay Accuracy, Study 1

	Distinctive Hearsay Accuracy			Distinctive Direct Accuracy			Hearsay Consensus		
	β	95% CI	*p*	*β*	95% CI	*p*	*r*	95% CI	*p*
A	.13	[-.06, .32]	.173	.22	[.05, .40]	.013	.20	[.02, .39]	.031
C	.06	[-.14, .26]	.562	.18	[-.00, .36]	.053	.36	[.20, .52]	<.001
HP	.06	[-.14, .27]	.543	.12	[-.07, .30]	.213	.34	[.18, .50]	<.001
N	.09	[-.10, .28]	.346	.15	[-.03, .32]	.109	.33	[.17, .49]	<.001
E	.07	[-.14, .28]	.521	.37	[.18, .56]	<.001	.58	[.45, .70]	<.001
O	.09	[-.10, .27]	.358	.30	[.14, .46]	<.001	.24	[.07, .41]	.006

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. *β* corresponds to the standardized regression slope; 95% CI corresponds to the 95% CI around the standardized regression slope.

Individual difference moderators. In total we examined seven potential moderators of hearsay consensus and hearsay accuracy: T, P1, and P2 extraversion, and P2 empathy across 4 subscales.

Extraversion. We assessed whether or not each member of the triad's extraversion moderated hearsay consensus and hearsay accuracy on each of the Big Six traits. In these analyses, we extended the analytic approach described earlier for hearsay consensus and hearsay accuracy by including the relevant extraversion variable as a moderator. For example, to analyze how Ts' extraversion moderated hearsay consensus for agreeableness, we regressed P1s' ratings of agreeableness onto P2s' ratings of agreeableness (mean-centered), Ts' extraversion (mean-centered), and the cross-product. Exchangeable estimates were pooled in the same way as other analyses. Similar analyses were carried out for all combinations of traits of interest with T, P1, and P2 extraversion, and for both hearsay consensus and hearsay accuracy.

Results from these analyses are presented in Supplemental Tables S4 and S5, but they can be summarized relatively briefly. The only significant interaction was targets' extraversion moderating hearsay accuracy for openness, β = .19, 95% CI [.02, .37], p = .032, suggesting that more Extraverted targets are more accurately perceived by P2s in terms of openness; (all other p's > .05). Given that this was only found for 1 of 36 analyses, and that its CI includes a negligible effect, it should be interpreted with some caution until it is replicated.

P2 empathy. We assessed the extent to which P2s' empathy, as measured by the four subscales of the IRI (Davis, 1983), moderated hearsay consensus and accuracy for the Big Six. Analyses were similar to those for extraversion. Results are displayed in Supplemental Table S6: going by significance, P2s' perspective taking moderated hearsay consensus for agreeableness (β = .19, 95% CI [.01, .37], p = .044) and openness (β = .19, 95% CI [.01, .37], p = .044). The rest of the interaction terms were not significant. Again, given the marginal p values and large number of tests (48 tests total) we approach interpreting these results with caution.

Discussion

Study 1 was initiated to test a new procedure and analytic framework to study hearsay reputations, to provide initial estimates of relative elevation, hearsay consensus, and hearsay accuracy, and to assess individual differences as moderators of hearsay accuracy and consensus.

We found differences in relative elevation, but the differences were not a simple fit to any single hypothesis. Evidence of self-enhancement was clearest for extraversion and openness and somewhat more ambiguous for other traits. The elevation data did not neatly fit the distance or hearsay-positivity hypotheses.

Hearsay consensus was moderately high across the board – indeed, on par with typical interjudge (i.e., P1-P1) consensus – and was especially high for extraversion. Hearsay accuracy

tended to be low across the board and was only distinguishable from chance guessing for extraversion. Likewise, follow-up analyses demonstrated that P2s did not achieve distinctive hearsay accuracy for any of the Big Six domains. Taken together, the flow of information from P1 to P2 appeared to be relatively strong, but the flow of information from T to P1 was relatively weaker; together, this led to P2s not receiving enough valid information to render an accurate judgement of the target. Finally, we found little evidence that extraversion (among any of the triad members) or empathy (among P2s) moderated hearsay consensus or accuracy.

Study 2

Study 1 demonstrated the viability of this new procedure and analytic framework for studying hearsay reputations and provided first estimates of important parameters. The estimates for hearsay consensus and accuracy were similar to meta-analytic benchmarks for reputations formed through direct interactions. For Study 2 we had three primary aims to build on these findings. First, we investigated how the context of the hearsay conversation affects hearsay consensus and accuracy, focusing on differences between status- and affiliation-oriented contexts. Second, we investigated how hearsay reputations of a target's personality may affect social decisions made about the target, similarly focusing on the status- and affiliation-oriented decisions of whether someone would make a good leader and friend. Third, Study 2 was a replication of Study 1 (with extensions), so we combined data from both studies wherever warranted in order to obtain more precise estimates.

For our first aim, we manipulated the context of the hearsay exchange by instructing participants to focus on different aspects of a target. For the *status* condition we instructed participants to discuss the target's leadership potential; for the *affiliation* condition we instructed them to discuss friendship potential; and for a *control* condition we instructed them to discuss

how the target is in general, similar to Study 1. We hypothesized that hearsay consensus and accuracy would be greater for extraversion in the status condition than in the control condition. Likewise, we hypothesized that hearsay consensus and accuracy would be greater for agreeableness in the affiliation condition than in the control condition. As described earlier, these hypotheses stem from two principles: conversational norms dictate that people focus on what is perceived to be relevant (Grice's maxim of relevance; Grice, 1975), and extraversion's and agreeableness's relevance to status and affiliation respectively (Anderson et al., 2001; Buss, 1996, 2011; Cottrell et al., 2007; Denissen & Penke, 2008; DeYoung, 2015; Judge et al., 2002; Lawless Desjardins et al., 2015; Wood, 2015; Wortman & Wood, 2011).

We additionally tested two hypotheses concerning the social consequences of information flowing through reputation networks. Specifically, we assessed the extent to which the information flowing from P1 to P2 about personality traits affected relevant social decisions P2 might make about the target. We hypothesized that information about the targets' extraversion flowing from P1 to P2 would affect the extent to which P2 thinks the target would make a good leader. We also hypothesized that information about the targets' agreeableness flowing from P1 to P2 would affect the extent to which P2 thinks the target would make a good friend.

Method

Pre-registration. As with Study 1, the measures, procedure, sample size, exclusion criteria, and analysis plan were pre-registered on OSF (https://www.osf.io/kfcmt). The full procedure (described below) was pre-registered and executed as described in the pre-registration. Our pre-registered sampling plan was to collect data until we had at least 120 targets (between 60 and 120 groups of 3 and 4 participants); we planned to continue collecting data beyond this point if it was minimally costly (e.g., continuing data collection until the end of the term). We had

three pre-registered exclusion criteria. First, groups were excluded if participants knew one or more of the people with whom they interacted (included in addendum to Study 1's pre-registration). Second, individual participants were excluded if they told one or more of the experimenters that they didn't understand the directions (inspired by Study 1). Third, the second author conducted a blinded screen (MacCoun & Perlmutter, 2015), wherein he inspected de-identified and unlinked datasets for invalid responding (as done in Study 1).

We present all of our primary pre-registered analyses. We additionally report several follow-up analyses; because the majority of analyses were pre-registered, we indicate which analyses were *not* pre-registered. Similar to Study 1, we had primary and secondary analyses (again, more properly first-order and second-order priorities). The primary analyses were tests of the hypotheses outlined above regarding the context manipulation and social consequences, as well as a replication of the Study 1 analyses in a combined dataset (the *Integrated* analyses). Like Study 1, we included these same results (i.e., hearsay consensus, etc.) on the additional trait-like items and Big Five facets as secondary (second-order priority) analyses; these aren't reported in the present manuscript to conserve space. Finally, the preregistration included a section of exploratory analyses; this is a potential misnomer, and likely should have been included as secondary analyses (since they are pre-specified analyses, just without strong predictions of results). The exploratory analyses included running the individual difference moderators (extraversion and P2 empathy[3]) in the combined data and looking at potential condition differences in hearsay consensus and accuracy for the remaining Big Six (and Big Five facets). Only the former (the individual difference moderators) are reported here. This study was approved by the University of Oregon's institutional review board.

[3] The pre-registration mistakenly says P1 or P2 empathy; we did not measure P1 empathy and never intended to measure it or run these analyses.

Participants. We recruited participants from the University of Oregon Human Subjects Pool. Participants signed up independently and were unaware of the topic of the study upon signing up. Our initial sample consisted of 381 participants in 108 groups. Groups of participants were excluded if any participant indicated that they knew one of the people they interacted with, which eliminated 3 groups (11 participants). One additional participant was excluded after they reported to the researcher(s) that they did not follow the questionnaire instructions. This resulted in a final sample of 369 participants in 105 groups (50 groups of 3; 55 groups of 4). We aimed to have at least 120 participants in the target role (across conditions); our final sample included 160 participants in the target role.

Participants were aged 18 to 32, with an average age of 19.4 (SD_{Age} = 1.8). The majority of participants identified as female (59% of those that reported). The majority of participants identified as White/Caucasian (62%), followed by Asian (14%), Hispanic/Latino (11%), Black/African American (5%), other (3%), Native American (2%); participants were able to select more than one option. 14% of participants did not complete the demographic questions.

Procedure. The procedure for Study 2 was nearly identical to Study 1 with one exception: groups were randomly assigned to one of three conditions (status, affiliation, or neutral/control). Once phase 2 (hearsay phase) of the experiment began, participants in the status and affiliation conditions were read the following instructions:

> We want you to talk about what that person is like in general. Later on, we are going to ask [P2] about [his or her] impressions of that person. *One thing we are going to ask is whether [P2] thinks the person you're describing would make a (good leader / good friend who they would want to get close to).* So we want you to help [P2] form an overall

impression*, which includes forming an impression of that person's (leadership potential / potential as a friend).*

In the control condition, the italicized sections above were removed. The full text of the instructions can be at the OSF project page (https://www.osf.io/65wfu). Outside of this and the additional measures administered after the interactions, the procedure was identical to Study 1.

Measures. All measures were pre-registered. Measures were identical to Study 1, with two additional questions administered in P1- and P2-reports (social decision items below).

Big Six. We administered self-, P1-, and P2-reports of the Big Six personality domains. We again used the full BFI-2 (Soto & John, 2017) and the 8-item honesty-propriety subscale from the 48QB6 (Thalmayer et al., 2011) for self-, P1-, and P2-reports. As in Study 1, all of these ratings were on a scale from 1 (disagree strongly) to 5 (agree strongly) with a neutral point of 3 (neutral; no opinion). As in Study 1, instructions were re-worded to appropriately reflect whom they were rating. Supplemental Table S1 contains coefficient alphas reflecting adequate internal consistency for all subscales.

Social decisions. We included two new additional items for P1- and P2-reports. Participants rated the extent to which they thought the target would be a good leader and good friend. These items were on the same 1 to 5 scale as other items administered.

Analytic approach. Data were analyzed using the same general approach outlined in Study 1, i.e., at the group-level with exchangeable estimates pooled using equality constraints. Unlike Study 1, Study 2 built on this approach by making extensive use of model comparisons and multi-group path analytic framework, both to test hypothesized differences among conditions (treating condition as grouping variable to test hypotheses about the context manipulation) and to integrate results across studies (treating study as grouping variable). All

model comparisons used χ^2 difference tests (χ^2_d), Akaike Information Criterion (AIC), and Bayesian Information Criteria (BIC) to assess relative model fit. We chose to use these three indices because they are differentially sensitive to parsimony: χ^2_d provides a traditional significance test and is not concerned with parsimony, AIC is a parsimony-adjusted fit measure which aims to choose the model predicted to perform best in new samples, and BIC is a parsimony-adjusted fit measure that aims to choose the true model and further assumes a true model will be relatively simple. We present χ^2_d (and the associated *p* value), difference in AIC (Δ_{AIC}), and difference in BIC (Δ_{BIC}). A significant χ^2_d indicates that the more constrained model fits the data significantly worse. For Δ_{AIC} and Δ_{BIC}, negative values indicate support for the more constrained model, and positive values indicate support for the less constrained model.

Assessing context differences. Four hypotheses concerned differences in hearsay consensus and accuracy across the randomly assigned experimental groups. As pre-registered, we modelled these differences using a multi-group path analytic framework, building on the path analytic framework used in Study 1. Specifically, we first modelled the data such that estimates could differ across conditions, and then assessed the decrement in fit incurred by constraining estimates hypothesized to differ (e.g., hearsay consensus for extraversion in the status and control conditions) to be equal to one another. Our hypotheses would be supported if such equality constraints worsened model fit (indicated by a significant χ^2_d test, a positive Δ_{AIC}, and a positive Δ_{BIC}).

Integrating across studies. As pre-registered, we conducted parallel analyses to those conducted for Study 1 on a dataset that combined the two studies, thus providing a single *integrated* estimate for each parameter of interest (e.g., hearsay consensus). In an additional analysis that was not pre-registered, we assessed the justifiability of integrating across studies

using a multi-group path analytic framework. We combined the data from both studies and estimated two models for each analysis: a *non-integrated model*, which allowed all model parameters to differ across studies, and an *integrated model*, which constrained all model parameters to be equal across studies. These were omnibus tests; they tested whether or not all parameters in the model could be pooled. For relative elevation and differential consensus/accuracy, this consisted of model comparisons testing whether the full variance-covariance and mean structure of the ratings could be pooled across studies, effectively testing whether all relative elevation and differential consensus/accuracy analyses could be integrated across studies in one test (for each Big Six domain). Since regression analyses (e.g., individual difference moderators) each required their own models, an analogous comparison was done for each regression model (i.e., testing whether all parameters in the regression model could be integrated across studies); those results will be presented before each corresponding analysis is presented.

Analyses were again run in the *lavaan* package (Version 0.6-3; Rosseel, 2012) and our own *ReputationModelR* package (see: https://www.github.com/coryc3133/ReputationModelR) in R (version 3.6.1; R Core Team, 2019). The data and a script that reproduces the analyses can be found at our OSF project page (https://www.osf.io/65wfu).

Results

Status-relevant context and extraversion. Our first hypothesis was that hearsay consensus for extraversion would be greater in the status condition than in the control condition. We tested this hypothesis by first estimating a model where hearsay consensus for extraversion could differ across conditions, then a model where hearsay consensus was equal in the status and control conditions. The hypothesis would be supported if the latter fit the data worse (as

determined by a significant χ^2_d and positive Δ_{AIC} and Δ_{BIC} values) and the results showed greater hearsay consensus for extraversion in the status condition.

Figure 5 displays hearsay consensus, hearsay accuracy, and direct accuracy for extraversion across condition. As seen in Figure 5, hearsay consensus for extraversion was slightly higher in the status condition (r = .68; 95% CI [.54, .82]) than the control condition (r = .57; 95% CI [.41, .74]) condition, but the model of no difference did not fit the data significantly worse (Δ_{AIC} = -0.65; Δ_{BIC} = -3.29; χ^2_d (1) = 1.35, p = .244). The negative Δ_{AIC}, Δ_{BIC}, and non-significant χ^2_d all suggest that hearsay consensus for extraversion does not differ across condition.

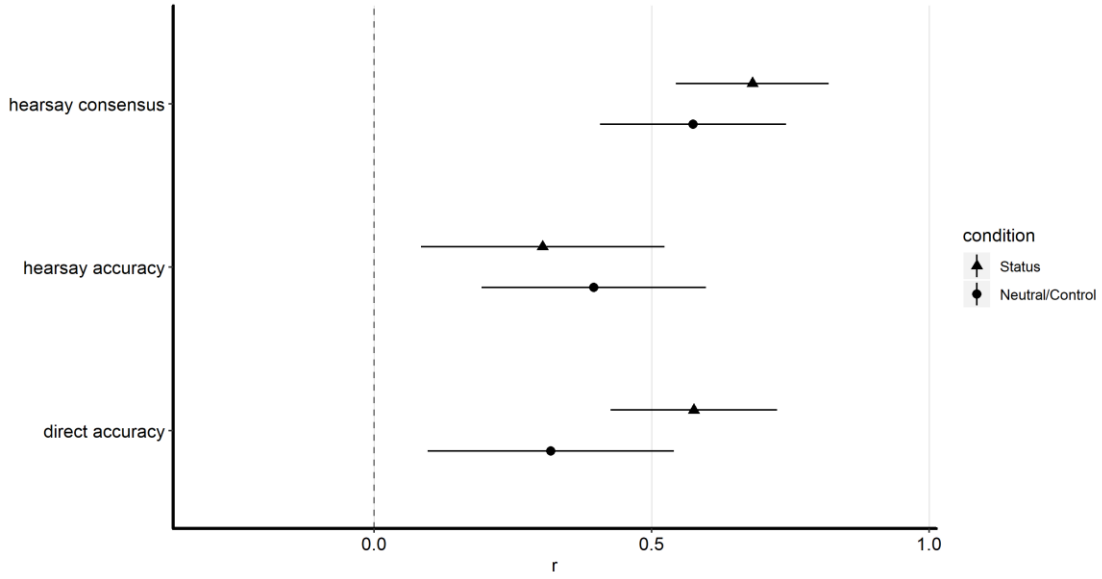


Figure 5. Hearsay consensus, accuracy, and direct accuracy for extraversion across conditions. Shapes represent point estimates and lines correspond to 95% CI's around those estimates.

Our second hypothesis predicted greater hearsay accuracy for extraversion in the status condition, and we tested it similarly. This hypothesis was also not supported. Hearsay accuracy for extraversion was slightly *lower* in the status condition (r = .30, 95% CI [.08, .52]) than the control condition (r = .40, 95% CI [.19, .60]). Moreover, the model consistent with no difference

did not fit the data worse according to χ^2_d and fit better according to Δ_{AIC} and Δ_{BIC} (Δ_{AIC} = -1.88; Δ_{BIC} = -4.52; χ^2_d (1) = 0.12, p = .727).

We thus found evidence mostly inconsistent with the idea that the status manipulation would affect the consistency or accuracy of hearsay reputations of extraversion. Follow-up analyses with direct accuracy for extraversion (the third parameter displayed in Figure 5) similarly showed no differences between the status and control conditions (see Supplemental Table S7 for model comparison results).

Affiliation-relevant context and agreeableness. We predicted that hearsay consensus for agreeableness would be greater in the affiliation condition than the control condition, which we tested similarly to the hypotheses for the status manipulation and extraversion. Figure 6 contains the estimates for hearsay consensus, hearsay accuracy, and direct accuracy for agreeableness across condition. Contrary to our hypothesis, we found a slightly lower estimate of hearsay consensus for agreeableness in the affiliation condition (r = .34, 95% CI [.10, .58]) than the control condition (r = .41, 95% CI [.19, .62]), with model comparison results consistent with no difference across these conditions (Δ_{AIC} = -1.89; Δ_{BIC} = -4.54; χ^2_d (1) = 0.11, p = .741). The hypothesis that hearsay accuracy for agreeableness would be greater in the affiliation condition was also not supported. Hearsay accuracy was descriptively higher in the affiliation condition (r = .26, 95% CI [.00, .53]) than the control condition (r = .03, 95% CI [.-.26, .32]) condition, but the model fit evidence was inconsistent with a difference (Δ_{AIC} = -1.06; Δ_{BIC} = -3.71; χ^2_d (1) = 0.94, p = .333).

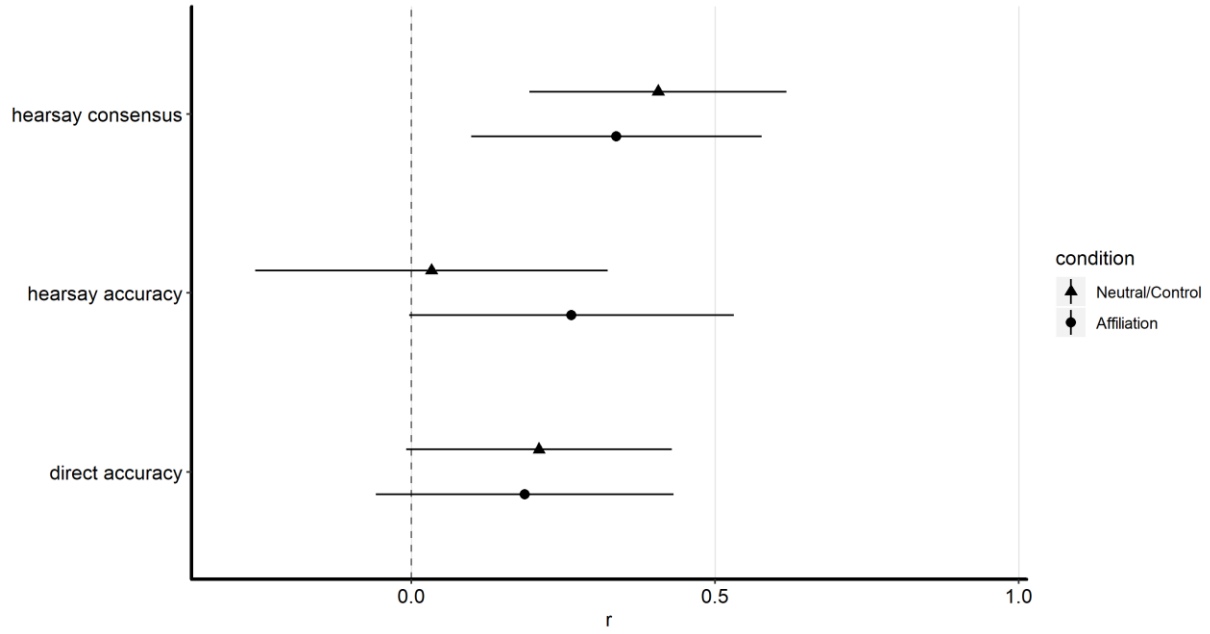
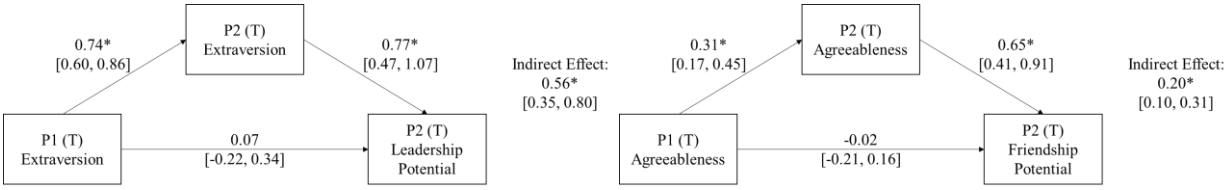


Figure 6. Hearsay consensus, accuracy, and direct accuracy for agreeableness across conditions. Shapes represent point estimates and lines correspond to 95% CI's around those estimates.

Thus, there was virtually no support for the hypothesis that the affiliation manipulation would increase hearsay consensus or accuracy for agreeableness. Follow-up analyses with direct accuracy (see Figure 6) similarly demonstrated no effect of the manipulation (see Supplemental Table S8 for model comparison results).

Effects of reputation on leadership judgments. We predicted that information about targets' extraversion flowing from P1s to P2s would affect P2s' perceptions of targets' leadership potential. We tested this hypothesis in a mediation model (with bootstrapped standard errors) that mirrored these predictions (see Figure 7a); like other analyses, this required pooling across exchangeable triads. Figure 7a shows this mediation model with the corresponding (unstandardized) regression slopes, 95% CIs, and *p* values. As seen in Figure 7a, we found the hypothesized relations: P1s' perceptions of targets' extraversion were shared with P2s ($b = 0.74$, 95% CI [0.60, 0.86], $p < .001$), which in turn predicted P2s' perceptions of targets' leadership

potential (b = 0.77, 95% CI [0.47, 1.07], p < .001). The indirect effect was significantly different from zero, moderately large in magnitude, and consistent with a moderate to large effect ($b_{indirect}$ = 0.56, 95% CI [0.35, 0.80], p < .001). In effect, this means that each scale point increase in P1s' perceptions of targets' extraversion is related to roughly half a scale point increase in P2s' perceptions of targets' leadership potential (with the 95% confidence interval ranging from one-third to four-fifths of a scale point increase), which appeared to be transmitted through P2s' perceptions of targets' extraversion.



(a) Extraversion and leadership judgments. (b) Agreeableness and friendship judgments.

Figures 7a-b. Mediation models. Standard errors were obtained via bootstrapping with 1000 resamples. The point estimate and 95% CI for the indirect effect are shown to the right. *p < .001

We next attempted to investigate the specificity of the extraversion-leadership connection and assess the possibility that the observed effect is driven by global positive evaluation, in a non-pre-registered follow-up analysis. We did so by regressing P2s' perceptions of targets' leadership potential on indirect effects for all of the Big Six in a multiple mediation model. A graphical depiction of this model is shown in Supplemental Figure S2. In this model, the indirect effect central of extraversion judgments remained significant, but did drop in magnitude ($b_{indirect}$ = 0.36, 95% CI [0.16, 0.58]), p = .001), now indicating a moderately small indirect effect and consistent with a small to moderate effect. Additionally, two non-hypothesized indirect effects emerged as significant. P1s' perceptions of the targets' conscientiousness were shared with P2s (b = 0.43, 95% CI [0.28, 0.59], p < .001), which in turn affected P2s' perceptions of targets' leadership potential (b = 0.60, 95% CI [0.23, 0.92], p = .001; b_{indirect} = 0.26, 95% CI [0.08, 0.46],

p = .009). Likewise, P1s' perceptions of targets' openness were shared with P2s (b = 0.19, 95% CI [0.08, 0.30], p < .001), which in turn affected their perceptions of targets' leadership potential (b = 0.61, 95% CI [0.24, 0.96], p = .002; $b_{indirect}$ = 0.12, 95% CI [0.03, 0.22], p = .016). Note that each of these indirect effects was smaller than the hypothesized effect, and each were consistent with negligible effect sizes (CIs ranged below 0.10). Taken together P1s' perceptions of targets' extraversion, conscientiousness, and openness influence P2s' perceptions of the targets' leadership potential such that a unit increase in each is expected to increase P2s' perceptions of targets' leadership potential by roughly one-third, one-fourth, and one-tenth of a likert-scale unit respectively.

Effects of reputation on friendship judgments. We also hypothesized that information about targets' agreeableness flowing from P1s to P2s would affect P2s' perceptions of targets' friendship potential. We tested this hypothesis via a mediation model which mirrored these predictions, pooling estimates from exchangeable triads. This model and the corresponding unstandardized regression slopes, 95% CIs, and p values are shown in Figure 7b. Consistent with H4, P1s shared information about targets' agreeableness with P2s (b = 0.31, 95% CI [0.17, 0.45], p < .001), which in turn affected P2s' perceptions of targets' friendship potential (b = 0.65, 95% CI [0.41, 0.91], p < .001; $b_{indirect}$ = 0.20, 95% CI [0.10, 0.32], p < .001; see Figure 7b). The indirect effect was significant and consistent with a small or moderate effect size. This means that each scale point increase in P1s' perceptions of a targets' agreeableness is associated with an increase of about one fifth of a scale point in P2s' perceptions of the targets' friendship potential, and this is achieved primarily via effecting P2s' perceptions of the targets' agreeableness.

Mirroring our approach for leadership potential, we conducted a non-pre-registered follow-up analysis aimed at assessing the specificity of the link between perceptions of

agreeableness and friendship potential by modelling hearsay reputations of friendship potential as a function of indirect effects of direct and hearsay reputations of each of the Big Six; this model is depicted in Supplemental Figure S3. The indirect effect of agreeableness dropped slightly but remained significant ($b_{indirect}$ = 0.15, 95% CI [0.05, 0.27], p = .008). The indirect effects for the remaining Big Six were small and not significant (all $|b_{indirect}|$'s < .05; all p's > .1). Thus, it appears that hearsay reputations of friendship potential were, within the Big Six, mostly a function of hearsay reputations (and indirectly, direct reputations) of agreeableness.

Integrated analyses. In order to obtain more precise estimates of relative elevation and differential consensus and accuracy, we planned a series of integrated analyses that combined data from Studies 1 and 2.

The first step was to assess the extent to which integrating across studies was justifiable, by comparing a *non-integrated model* (a model where parameters can be different for each study) to the *integrated model* (a model where parameters are the same for both studies). All of the relative elevation and differential consensus/accuracy were tested in a single omnibus model for each Big Six domain. The results of these models are shown in Supplemental Table S9. The χ_d^2 test was not significant for any of the domains (p's $\geq$ .05), both AIC and BIC were lower (better) for the integrated model for all trait domains (Δ_{AIC}'s from -5.04 to -20.50; Δ_{BIC}'s from -52.68 to -68.14). This suggests that the allowing studies to differ does not increase fit enough to justify the added complexity of the model, and so lends support to integrating across studies. Thus, the evidence unequivocally justified integrating across studies for all Big Six domains. For the sake of completeness, the integrated *and* (non-integrated) Study 2 results will be included in each table and figure, but only the integrated results will be presented in text.

Relative elevation. Relative elevation was assessed identically to Study 1, with the addition of pooling across studies. The pooled means and SDs used for these analyses can be found in Supplemental Table S10, and the results of the comparison in terms of a raw difference (in likert-scale units), a 95% CI around the raw difference, a Cohen's *d*, and associated *p* values can be found in Table 3. As in Study 1, the raw differences should be interpreted in terms of a one to five Likert scale.

Table 3

Relative Elevation Differences, Integrated and Study 2 Analyses

		Integrated				Study 2			
		D	95% CI	*d*	*p*	D	95% CI	*d*	*p*
P1-P2	A	0.20	[0.13, 0.28]	0.37	<.001	0.13	[0.03, 0.23]	0.24	.008
	C	0.03	[-0.04, 0.11]	0.06	.390	-0.05	[-0.14, 0.05]	-0.07	.362
	HP	0.26	[0.18, 0.33]	0.46	<.001	0.16	[0.06, 0.26]	0.29	.002
	N	-0.14	[-0.20, -0.07]	-0.25	<.001	-0.08	[-0.17, 0.01]	-0.14	.066
	E	-0.04	[-0.11, 0.04]	-0.05	.349	-0.03	[-0.13, 0.06]	-0.04	.490
	O	0.11	[0.05, 0.18]	0.23	.001	0.09	[0.00, 0.18]	0.17	.050
Self-P2	A	0.15	[0.07, 0.23]	0.28	<.001	0.13	[0.03, 0.24]	0.25	.010
	C	-0.14	[-0.23, -0.05]	-0.22	.003	-0.2	[-0.32, -0.08]	-0.31	.001
	HP	0.20	[0.12, 0.28]	0.35	<.001	0.14	[0.03, 0.25]	0.25	.012
	N	0.19	[0.09, 0.29]	0.27	<.001	0.21	[0.09, 0.34]	0.32	.001
	E	0.14	[0.05, 0.24]	0.19	.003	0.12	[0.00, 0.23]	0.15	.050
	O	0.47	[0.40, 0.54]	0.92	<.001	0.45	[0.37, 0.54]	0.92	<.001
Self-P1	A	-0.05	[-0.11, 0.01]	-0.09	.130	0.00	[-0.07, 0.08]	0.00	.984
	C	-0.17	[-0.24, -0.10]	-0.28	<.001	-0.15	[-0.24, -0.06]	-0.24	.001
	HP	-0.06	[-0.13, 0.01]	-0.11	.096	-0.02	[-0.11, 0.06]	-0.04	.588
	N	0.33	[0.24, 0.42]	0.47	<.001	0.3	[0.19, 0.40]	0.42	<.001
	E	0.18	[0.10, 0.26]	0.25	<.001	0.15	[0.04, 0.25]	0.21	.005
	O	0.36	[0.29, 0.42]	0.65	<.001	0.37	[0.28, 0.45]	0.65	<.001

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. P1-P2 = P1(T) – P2(T); Self-P2 = T(T) – P2(T); Self-P1 = T(T) – P1(T). D denotes raw differences; *d* denotes Cohen's *d* values. 95% CIs correspond to raw differences. Values are obtained from a model pooling means and variances for exchangeable triad members. S2 = Study 2 estimate; Pooled = Estimate pooled across both studies.

P1-P2 relative elevation. How much did direct and hearsay reputations differ on average? As seen in left side of Table 3, P1s rated the target as more agreeable by approximately one third of a SD, more honest/proper by almost half a SD, less neurotic by a quarter of a SD, and more open by almost a quarter of a SD. The narrower confidence bands afforded by pooling the two samples indicate that these results are consistent with small-to-moderate differences, with

neuroticism and openness including negligible differences (-0.07 and 0.04 in raw scale units respectively) at the lower bound. The pooled analyses thus further suggest that direct reputations tend to be at moderately more positive than hearsay reputations, with major exceptions in conscientiousness and extraversion where they were equal.

Self-P2 relative elevation. How much did hearsay reputations differ from self-perceptions? Self-P2 relative elevation results are shown in Table 3. P2s rated targets as less agreeable by approximately a quarter of an SD, more conscientious by about one-fifth of a SD, less honest by a third of a SD, less neurotic by just over a quarter of a SD, less extraverted by almost one-fifth of a SD, and less open by almost a full SD. Self-P2 relative elevation for openness was quite high in magnitude, with the lower limit of its CI greater than the upper limit of any of the other five domains. Thus, the pooled results lend further evidence to hearsay reputations being less positive than self-knowledge, with the exceptions of conscientiousness and neuroticism, where hearsay reputations were more favorable than self-knowledge.

Self-P1 relative elevation. How different were direct reputations and self-perceptions? Put differently, did we see evidence of self-enhancement or self-effacement? Self-P1 relative elevation results are in Table 3. Compared to self-ratings, P1s rated targets as more conscientious by about a fifth of a SD, less neurotic by almost half of a SD, less extraverted by about a quarter of a SD, and less open by about two-thirds of a SD. Unlike analyses with just the Study 1 data, the more precise pooled analyses suggest virtually no difference in agreeableness or honesty-propriety (see Table 3). The CIs around the raw estimates were consistent with moderate-to-large differences in neuroticism and openness and small-to-moderate differences in conscientiousness and extraversion. Thus, direct reputations appeared to be less positive than self-knowledge with respect to extraversion and openness and more positive with respect to conscientiousness and

neuroticism, which partly map onto the Big Two domains of Dynamism and Social Self-Regulation (Saucier & Srivastava, 2015).

Differential consensus and accuracy. Differential consensus and accuracy pertain to the rank-ordering of ratings. As above, we emphasize the integrated results in the text, but we report Study 2 results in figures for completeness.

Hearsay consensus. As in Study 1, hearsay consensus was assessed by correlating P1s' ratings of the targets with P2s' ratings of the targets separately for each of the Big Six. These correlations speak to the strength (but not necessarily accuracy) of information about targets flowing from P1s to P2s. Figure 8 presents hearsay consensus estimated in each study separately and combined (i.e., integrated), alongside the same meta-analytic benchmark used in Study 1 (i.e., interjudge consensus for previously unacquainted participants; Connelly & Ones, 2010). As seen in Figure 8, hearsay consensus was moderately high for agreeableness (r = .28, 95% CI [.17, .39], p < .001), conscientiousness (r = .39, 95% CI [.29, .49] p < .001), honesty-propriety (r = .29, 95% CI [.18, .40] p < .001), neuroticism (r = .34, 95% CI [.24, .44] p < .001), quite high for extraversion (r = .62, 95% CI [.55, .69] p < .001), and moderately high for openness (r = .25, 95% CI [.15, .36] p < .001).

In summary, hearsay consensus was consistent across domains, and similar to meta-analytic estimates of interjudge consensus. The exception was extraversion, where hearsay consensus was again higher than the other domains and the meta-analytic estimate of interjudge consensus for extraversion. Indeed, the 95% CI around the integrated estimate for extraversion neither overlapped with other domains' CIs nor included the meta-analytic benchmark. Thus, hearsay leads to consensus generally, but consensus for extraversion at levels that exceed other domains, or typical (P1-P1) interjudge consensus.

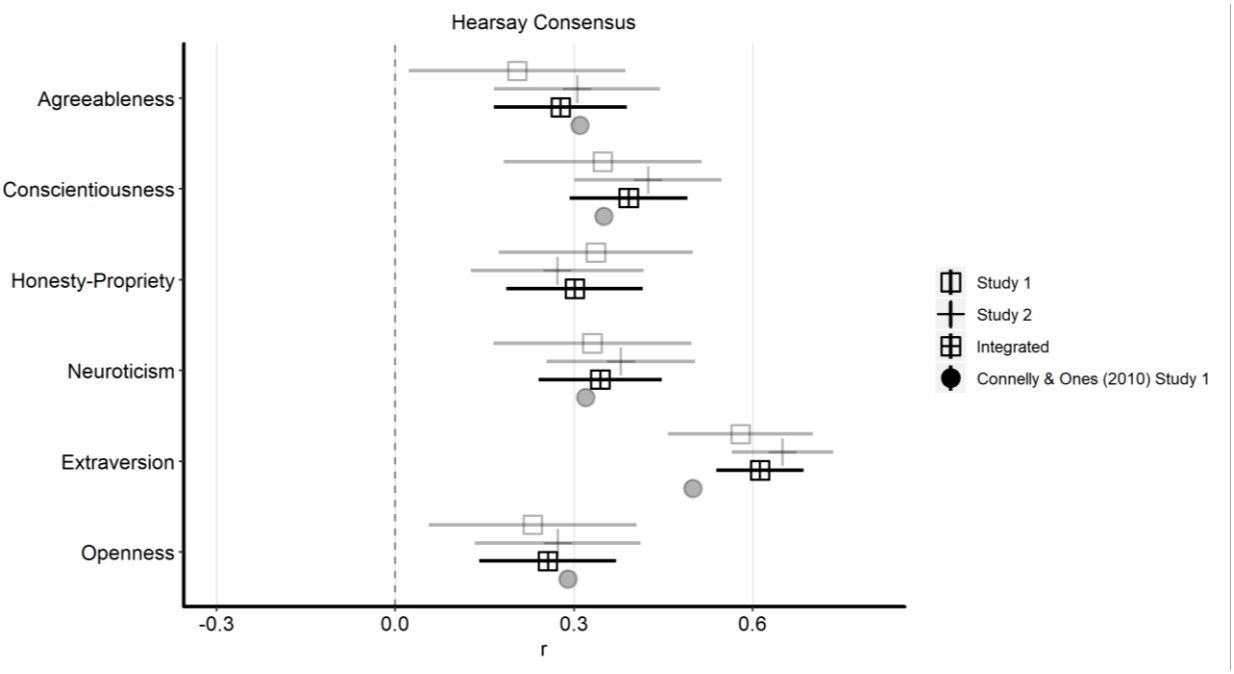


Figure 8. Study 1, Study 2, and integrated estimates of hearsay consensus for the Big Six. Study 1 (box) estimates are based just on the Study 1 sample; Study 2 (cross) estimates are based just on the Study 2 sample (collapsed across conditions); integrated estimates (crossed-box) are based on data from both Studies. Circles again represent meta-analytic estimates of interjudge consensus following naturalistic laboratory interactions, obtained from Study 1 of Connelly & Ones's (2010) meta-analysis on consensus and self-other agreement.

Hearsay accuracy. Like Study 1, hearsay accuracy was assessed by correlating P2s' ratings of the targets with the targets' self-reports separately for each of the Big Six. It speaks to the extent to which the information flowing from targets to P1s to P2 is valid or accurate. Figure 9 contains the integrated estimates of hearsay accuracy for the Big Six alongside each study's (non-integrated) estimate and the meta-analytic benchmark of self-other agreement among previously unacquainted participants (Connelly & Ones, 2010). As seen in Figure 9, hearsay accuracy was low and indistinguishable from chance guessing for agreeableness (r = .09, 95% CI [-.03, .21], p = .151), moderately low but distinguishable from chance guessing for conscientiousness (r = .16, 95% CI [.04, .27], p = .007), low and indistinguishable from chance guessing for honesty-propriety (r = .06, 95% CI [-.07, .18], p =.372), moderately low but distinguishable from chance guessing for neuroticism (r = .15, 95% CI [.03, .26], p = .012), and

moderately high and distinguishable from chance guessing for extraversion (r = .31, 95% CI [.21, .41], p < .001) and openness (r = .24, 95% CI [.12, .35], p < .001).

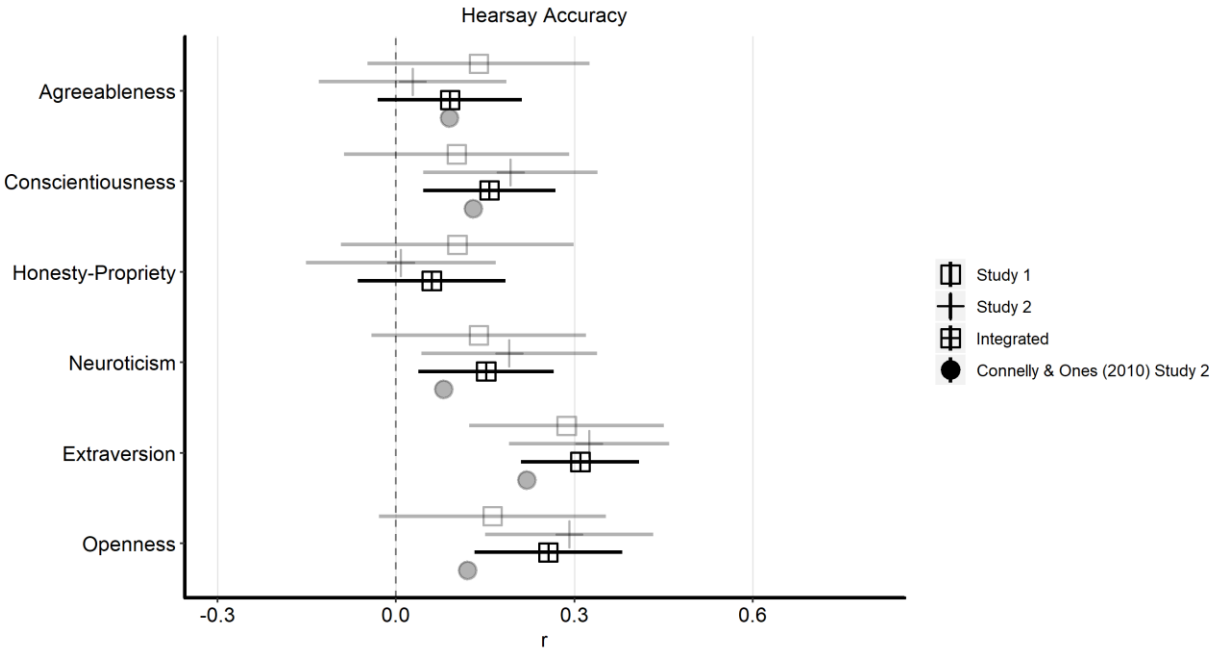


Figure 9. Study 1, Study 2, and integrated estimates of hearsay accuracy for the Big Six. Study 1 (box) estimates are based just on the Study 1 sample; Study 2 (cross) estimates are based just on the Study 2 sample (ignoring condition); integrated estimates (crossed-box) are based on data from both Studies. Circles represent meta-analytic estimates of self-other agreement among previously unacquainted participants, obtained from Study 2 of Connelly & Ones's (2010) meta-analysis on consensus and self-other agreement (See rows labelled "Stranger" in Table 5 in Connelly & Ones, 2010).

Descriptively, Figure 9 shows variability across Big Six domains. Hearsay accuracy appears to be moderately high for extraversion and openness, moderately low but not zero for conscientiousness and neuroticism, and indistinguishable from zero for agreeableness and honesty-propriety. Finally, CIs for hearsay accuracy overlapped with meta-analytic benchmarks of self-other agreement from direct reputations, suggesting we also cannot rule out the possibility that P2s reach similar degrees of accuracy as meta-analytic estimates suggest P1s do.

Direct accuracy. Direct accuracy (i.e., self-P1 agreement) was highly consistent across studies; the integrated estimates were similar to meta-analytic estimates (Connelly & Ones, 2010), but tended to be a little bit higher in our study (see Supplemental Figure S4).

Distinctive hearsay accuracy. Parallel to Study 1, we conducted a non-pre-registered follow-up analysis on the extent to which there was distinctive hearsay accuracy. This analyses addressed whether P2s knew something accurate about the target that P1s did not. We assessed this in the same manner as Study 1. As with all integrated analyses, we first assessed whether integrating across studies led to a detriment in fit, and these model comparison results can be found in Supplemental Table S11. We found unambiguous support for integrating across studies, as demonstrated by non-significant χ_d^2's (all *p*'s > .05), negative Δ_{AIC} (from -2.84 to -14.18) and Δ_{BIC} (from -31.43 to -42.77) values (see Supplemental Table S11).

The integrated model parameters for distinctive hearsay accuracy are shown in Table 4. As shown in Table 4, there was some evidence of distinctive hearsay accuracy for openness (β = .17, 95% CI [.06, .29], *p* = .004), but not for the other Big Six (all β's ≤ .1; *p*'s > .1; see Table 4). These results indicate that P2s primarily achieve accuracy by adopting P1s view of the target, with the possible exception of openness, where P2s appear to achieve accuracy above and beyond what they achieve through agreement with P1s.

Table 4

Distinctive Hearsay Accuracy, Integrated and Study 2 Analyses

		Distinctive Hearsay Accuracy			Distinctive Direct Accuracy			Hearsay Consensus		
		β	95% CI	*p*	β	95% CI	*p*	*r*	95% CI	*p*
Integrated	A	.03	[-.09, .15]	.657	.24	[.13, .34]	<.001	.28	[.17, .39]	<.001
	C	.06	[-.06, .19]	.328	.27	[.16, .37]	<.001	.40	[.30, .50]	<.001
	HP	.02	[-.11, .14]	.780	.16	[.05, .27]	.005	.29	[.18, .39]	<.001
	N	.10	[-.03, .22]	.119	.18	[.07, .29]	.002	.35	[.25, .45]	<.001
	E	.11	[-.03, .26]	.113	.32	[.18, .45]	<.001	.62	[.55, .69]	<.001
	O	.17	[.06, .29]	.004	.23	[.13, .34]	<.001	.25	[.15, .36]	<.001

Study 2	A	-.05	[-.20, .11]	.535	.26	[.13, .40]	<.001	.32	[.19, .46]	<.001
	C	.05	[-.11, .21]	.564	.33	[.19, .47]	<.001	.42	[.29, .54]	<.001
	HP	-.02	[-.18, .15]	.850	.19	[.04, .33]	.011	.27	[.12, .41]	<.001
	N	.11	[-.06, .27]	.195	.20	[.05, .34]	.007	.38	[.26, .50]	<.001
	E	.15	[-.04, .34]	.111	.27	[.09, .45]	.003	.65	[.56, .73]	<.001
	O	.23	[.08, .38]	.002	.20	[.06, .33]	.005	.27	[.13, .41]	<.001

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. β corresponds to the standardized regression slope; 95% CI corresponds to the 95% CI around the standardized regression slope.

Individual difference moderators. As in Study 1, we assessed the extent to which targets', P1s', and P2s' extraversion and P2s' empathy moderated hearsay consensus and accuracy. These were examined in the same way as Study 1, but in a multi-group path analytic framework testing for between study differences like the other integrated analyses.

Extraversion. We examined the extent to which targets', P1s', and P2s'own (self-reported) extraversion moderated hearsay consensus and accuracy for each of the Big Six. As with other integrated analyses, we first examined the justifiability of integrating across studies. This resulted in 35 model comparisons (3 moderators X 6 domains X 2 outcomes minus 1 since target extraversion moderating hearsay accuracy isn't analyzable). The results of these 35 comparisons are shown in Supplemental Table S12 and S13, which contain comparisons for hearsay consensus and accuracy respectively. With just a few exceptions, model comparison results provided unequivocal evidence for integrating across studies.

We summarize the findings here; full results of extraversion moderating hearsay consensus using the integrated and just the Study 2 data can be found in Supplemental Tables S14 and S15 respectively. Target extraversion moderated hearsay consensus for agreeableness (β = .16, 95% CI [.05, .28], p = .005), meaning that triads with more Extraverted targets had greater hearsay consensus (P1-P2 agreement) for agreeableness. P1 extraversion moderated hearsay consensus for extraversion (β = .10, 95% CI [.01, .20], p = .005), suggesting that triads with

more extraverted P1s had slightly greater hearsay consensus for targets' extraversion. P2 extraversion moderated hearsay consensus for agreeableness (β = -.13, 95% CI [-.24, -.01], p = .031) and conscientiousness (β = -.12, 95% CI [-.23, -.01], p = .031), suggesting that triads with more Extraverted P2s had slightly *less* hearsay consensus for targets' agreeableness and conscientiousness.

Moderator analyses for hearsay accuracy produced two significant results; full results of extraversion moderating hearsay accuracy using the integrated and just the Study 2 data can be found in Supplemental Tables S16 and S17 respectively. P1 extraversion moderated hearsay accuracy for extraversion (β = .19, 95% CI [.16, .23], p < .001), suggesting that more extraverted P1s communicated a more accurate impression of the target to P2. P2 extraversion moderated hearsay accuracy for extraversion (β = -.13, 95% CI [-.25, -.02], p = .019), suggesting that more extraverted P2s are less accurate in their perception of targets' extraversion. All other interactions were not significant in the integrated data.

Given the large number of tests, small effect sizes, and relatively large p values (p's ≥ .005), we concluded that there was very modest evidence that Target, P1, and P2 Extraversion moderated hearsay consensus or accuracy.

P2 empathy. As with other integrated analyses, we first assessed the justifiability of integrating across studies; the results of comparing the integrated and non-integrated models for these moderator analyses are in Supplemental Table S18. This resulted in 48 model comparisons (4 subscales X 6 personality domains X 2 for hearsay consensus and accuracy). The results of these 48 comparisons are shown in Supplemental Table S18. With just a few exceptions, model comparison results provided unequivocal evidence for integrating across studies.

We summarize the empathy moderator analyses here; full results are in Supplemental Tables S19 and S20, which show results with the integrated and just Study 2 data respectively. Empathic concern appeared to moderate hearsay consensus for agreeableness (β = -.14, 95% CI [-.26, -.02], p = .021), conscientiousness (β = -.15, 95% CI [-.25, -.04], p = .007), and honesty-propriety (β = -.15, 95% CI [-.26, -.03], p = .015), suggesting that P2s with higher empathic concern reached lower degrees of hearsay consensus in these three domains. The remaining 21 P2 empathy moderating hearsay consensus analyses were not significant. Turning to hearsay accuracy, perspective taking appeared to moderate hearsay accuracy for honesty-propriety, β = .17, 95% CI [.05, .30], p = .007, suggesting that P2s with greater levels of perspective taking were more accurate in their perception of the targets' honesty-propriety. The remaining 23 empathy moderating hearsay accuracy analyses were not significant. Taken together, there was relatively modest evidence that P2s' empathy moderated hearsay consensus or accuracy, with a few possible exceptions.

General Discussion

Across two studies, we used the network approach to reputations to document how information flows across networks of perceivers in a naturalistic laboratory setting. Our findings revealed a tendency for second-degree perceivers (P2s) to form more negative impressions than people who directly interacted with a target (P1s), with some exceptions. Hearsay consensus was substantial across the board, and especially high for extraversion, suggesting that information can flow robustly between people discussing a third person who is not present. Hearsay accuracy was lower than consensus, and it was similar to estimates of direct accuracy from the literature and the present studies. Hearsay consensus and accuracy were not moderated by extraversion or empathy, nor were they moderated by a context manipulation giving participants a

conversational goal in the hearsay exchange. We found that the transmission of reputational knowledge can have consequences for important social decisions people make about one another: P1s shared information about targets' Big Six characteristics with P2s, which in turn appeared to influence the extent to which P2s thought the targets would be good leaders and friends.

We believe these results demonstrate the usefulness of the network approach to reputation, but they represent merely a first step in a broader research program. The present findings point to several promising avenues for this research program's next steps.

Documenting the Flow of Reputational Information

The network approach to reputations conceptualizes reputations as information flowing across a network of perceivers. By comparing the information stored in the nodes of that network – that is, the impressions formed by perceivers at different distances from a target – we can draw inferences about the valence, strength, and validity of information flow between people. The present findings have implications for each.

Relative elevation and valence. In comparisons of average ratings made by participants in different parts of the network, we found a number of interesting differences in relative elevation, but no single pattern that generalized across traits. Evidence for self-enhancement was clearest for extraversion and openness, where the integrated analyses showed that self-ratings were more positive than ratings made by both P1s and P2s. In higher-order factor models, extraversion and openness form a dynamism factor that reflects agentic engagement with the world and concerns with getting ahead (Saucier & Srivastava, 2015) and in self-enhancement terms reflects egoistic biases (Paulhus & John, 1998). We did not observe a consistent pattern of self-enhancement for the other traits, which are part of a higher-order social self-regulation factor

that reflects communal concerns for getting along and in self-enhancement corresponds to moralistic biases.

We also considered two more speculative hypotheses about relative elevation. One was that positivity would decrease as a function of distance from a target person: only openness was consistent with this hypothesis. The second was that norms against spreading negative gossip would lead to P2s forming positively biased impressions relative to P1s. For four traits, the reverse of this hypothesis held: P1s had more positive impressions than P2s of targets' agreeableness, honesty-propriety, neuroticism, and openness (differences for conscientiousness and extraversion were negligible). Given that the elevation results included some substantial effects but did not unambiguously support any one hypothesis across most or all trait domains, it is possible that different processes, including ones that we may not have considered here, may be operating for different traits. The one thing we are relatively confident of is that the relative elevation results cannot be adequately captured by a single, simple explanation.

Although P1s' impressions were more positively valanced than P2s' for most traits, it is possible that this is driven by tie strength (i.e., acquaintanceship) rather than positionality (P1 vs. P2) per se, since P2s never met P1s. Future work could tease this apart by comparing elevation of reputations between P1s who have different levels of acquaintanceship with targets. Outside of the experimental setting we created for this study, multiplexity – being associated with the target in multiple networks – may also be relevant, especially for certain kinds of other networks (friend networks, kin networks, etc.). Future work will need to disentangle the effects of positionality, tie strength, and multiplexity on the valence of reputational information. The network approach to reputations affords the theoretical, methodological, and analytic foundation to begin posing and answering such questions.

Differential consensus and accuracy. The hearsay consensus results showed substantial agreement between P1s and P2s. Hearsay thus appeared to contain a consistent signal. This agreement sometimes exceeded meta-analytic estimates of P1-P1 agreement, such as for extraversion. The signal appeared to be only somewhat valid, however, as demonstrated by the more modest hearsay accuracy estimates. The information flowing from T to P1 has been the primary focus of prior research on interpersonal perception and tends to be moderate (see Connelly & Ones, 2010). If anything, that link in the information flow was stronger in the present studies than a typical study among previously unacquainted participants. So why did we find little hearsay accuracy?

One possibility is that most of the signal loss occurs between T and P1, but that P1s are then reasonably effective at sharing almost all of the valid information they have. This would be broadly consistent with our results – estimates of direct and hearsay accuracy tended to be more similar to one another than either were to estimates of hearsay consensus. This would also be generally consistent with prior work demonstrating that P1-P1 consensus generally exceeds self-P1 agreement (Connelly & Ones, 2010), and with theoretical models emphasizing the unique perspective of the self (Vazire, 2010).

A second, potentially complementary possibility draws upon insights from Kenny's (2004) PERSON model of interpersonal perception. As outlined in the introduction, PERSON breaks interpersonal information down into different components. This includes stable information that is shared by all perceivers (the stereotype and personality components) and idiosyncratic but stable information that is unique to a single perceiver (the opinion component). PERSON suggests that any individual P1's impression is dominated by opinion. However, individual perceivers cannot distinguish between stereotype, personality, and opinion in their

own impression (the components can only be distinguished statistically). As a result, when perceivers share information about a target, they would share all three indiscriminately. This would lead to a strong, consistent, but substantially invalid, signal being communicated from P1s to P2s. Future work could examine this possibility by examining consensus and accuracy of reputations in larger networks. The stereotype and personality components originate from a target and therefore would spread through the entire network, producing global consistency among all nodes. By contrast, an opinion component originates with one perceiver, and therefore would spread only to other nodes who received information from that perceiver, producing local consistency in only a part of a network.

The amount of hearsay consensus and accuracy was different for the different Big Six domains, suggesting that hearsay may have more signal for some domains (e.g., extraversion) than others (e.g., honesty-propriety). Indeed, hearsay consensus and accuracy followed similar patterns to meta-analytic estimates of direct consensus (P1-P1) and accuracy (self-P1). The comparisons across traits roughly tracked how observable different traits are, suggesting that P1s might follow Grice's (1975) maxim of quality: P1s may disproportionately share information about Ts' standing on traits they feel they have some knowledge about (e.g., extraversion). Future work could test this more directly by manipulating which trait domains P1s know the most about, for example through a T-P1 context manipulation.

What Moderates the Flow of Reputational Information?

The individual differences we examined appeared to have little or no moderating effect on the extent of hearsay consensus and accuracy. Likewise, the context of the hearsay exchange which we experimentally manipulated in Study 2 had virtually no impact as well. We thus found

little evidence concerning moderators of the flow of reputational information, despite probing for such moderators in different ways.

Individual differences in targets' extraversion have been previously found to moderate direct accuracy (self-P1 agreement). Extraverts are more behaviorally expressive, leading to more cues available for perceivers to use in their judgment, ultimately enhancing perceivers' accuracy (Funder, 1995; Human & Biesanz, 2013). The lack of moderation by targets' extraversion was therefore somewhat surprising. It is possible that any impact targets' extraversion has on the accuracy of P1s judgment is too weak to have a detectable impact on P2s' judgments. Put differently, Extraverted Ts may send more valid signal to P1s, but that small increase in valid signal may be swamped by the invalid (but consistent) signal sent from P1s to P2s. Target effects on accuracy might be weaker for hearsay reputations (compared to direct reputations) more generally, which should be investigated more fully in future work.

More speculatively, we hypothesized that P1s and P2s higher in extraversion would achieve greater hearsay consensus and accuracy, by providing or eliciting more information about targets. We also speculated that P2s higher in empathy would draw more accurate inferences. We found no support for any of these hypotheses.

In Study 2 we used an experimental manipulation to focus the hearsay exchange on specific aspects of the target – the extent to which they would make a good leader or friend. The manipulation had little impact on the resulting impressions. There are several possible explanations for this finding. First, it may be that relevancy has little or no impact on hearsay; perhaps people share information they have about a target person regardless of how relevant that information is to P2. We believe this broad conclusion is unlikely, and more work would be needed to make a strong case that relevancy has no impact on reputational discourse.

Another possibility is that the instruction manipulation used in Study 2 may have been too weak. Previous work has demonstrated an effect of subtle instruction manipulations on direct accuracy (Biesanz & Human, 2010), but perhaps a stronger manipulation is needed to affect hearsay consensus or accuracy. Future work could consider increasing the incentive of a consistent or accurate impression, perhaps by having some perceived consequence for P1s or P2s. For example, P2s could be incentivized to achieve accurate perceptions of Ts by telling them that they will get to choose between the person they're hearing about and some other (unknown) person for particular tasks relevant to different personality domains. Attaching P2s' impressions to a perceived consequence might increase the impact of differences in relevancy on hearsay consensus and accuracy.

There are of course many additional factors that might influence hearsay reputations that we did not systematically examine in either of the above studies. Three important dimensions are tie strength among network members (i.e., both T-P1 and P1-P2 tie strength), multiplexity among network members (i.e., whether or not T-P1 and P1-P2 are additionally tied in friendship, kinship, or other networks), and any goals each person has with respect to the targets' reputation (e.g., if P1 wants to present T positively). It's possible that each of these may impact the flow of reputational knowledge on their own or through their interaction(s). Investigating how these and other dimensions affect the flow of reputational information in the laboratory and in real reputation networks is an important priority for future work.

Consequences of Hearsay Reputations

In Study 2, we attempted to document how reputations could affect social decisions that hearsay perceivers make about the target. As hypothesized, we found that information flowed from P1s to P2s about targets' standing on extraversion and agreeableness, which in turn affected

P2s' perceptions of targets' leadership and friendship potential respectively. We additionally found that information about targets' standing on conscientiousness and openness likewise flowed from P1 to P2 and affected P2s' perception of targets' leadership potential. These results demonstrate that hearsay reputations not only include information about targets' personality but additionally provide P2s with information they might use in deciding if and how to interact with the target person(s). Thus, the information that flows about a person – their direct and hearsay reputations – might influence the opportunities one is afforded in love and work.

The social consequences we examined here were ratings made by perceivers, and future work will be needed to establish real impacts on behavior. For example, future work could examine if hearsay reputations affect whom is selected for a task in a naturalistic laboratory study, or research with real reputation networks could examine how hearsay reputations affect the formation and dissolution of friendships and romantic relationships. Work amongst cohorts of new hires in organizational settings could examine how reputational information spreads in newly formed groups, and whether it shapes the opportunities one receives at work. In either case, the present work suggests that information can be shared from P1s to P2s with relative ease, and at least hints that some of this information can be quite consequential. Examining these processes in real networks on real consequences in individuals' real lives is an important step for future work.

Future Directions

Craik's (2009) network conceptualization, combined with the concepts and tools of interpersonal perception research, offers a powerful new way to theorize about and empirically study reputations. Inspired by that conceptualization, we developed an experimental paradigm and a data-analytic framework to study how reputations form and spread through micro-

networks. In two studies, we applied this approach to studying the spread of reputational information about personality among previously-unacquainted participants. But we see these initial studies and results as only a beginning, and we have made our materials and an R package available to facilitate adoption of this paradigm by other researchers. Real-world reputations can contain a wide range of information; and they spread through larger and more complex networks, among perceivers in many different contexts and with many different goals, through people with many different kinds of relationships. It is our hope that other researchers will see the importance and potential of this work, and that they will be as excited as we are to continue it in the future.

References

Allport, G. W., & Odbert, H. S. (1936). Trait-Names: A Psycho-lexical Study. *Psychological Monographs*, *47*, i–171. http://dx.doi.org/10.1037/h0093360

Anderson, C., John, O. P., Keltner, D., & Kring, A. M. (2001). Who attains social status? Effects of personality and physical attractiveness in social groups. *Journal of Personality & Social Psychology*, *81*, 116–132. https://doi.org/10.1037//0022-3514.81.1.116

Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The Experimental Generation of Interpersonal Closeness: A Procedure and Some Preliminary Findings. *Personality and Social Psychology Bulletin*, *23*, 363–377. https://doi.org/10.1177/0146167297234003

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, *21*, 372–374. https://doi.org/10.1177/0956797609360756

Biesanz, J. C. (2010). The Social Accuracy Model of Interpersonal Perception: Assessing Individual Differences in Perceptive and Expressive Accuracy. *Multivariate Behavioral Research*, *45*, 853–885. https://doi.org/10.1080/00273171.2010.519262

Biesanz, J. C., & Human, L. J. (2010). The cost of forming more accurate impressions: accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. *Psychological Science*, *21*, 589–594. https://doi.org/10.1177/0956797610364121

Buss, D. M. (2011). Personality and the adaptive landscape. In D. M. Buss & P. H. Hawley (Eds.), *The Evolution of Personality and Individual Differences* (pp. 29-57).

Buss, D. M. (1996). Social adaptation and five major factors of personality. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical Perspectives* (pp. 180–207). New York: Guilford Press.

Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-Insight: Do People Really Know How Others See Them? *Journal of Personality & Social Psychology*, *101*, 831–846. https://doi.org/10.1037/a0024297

Chaplin, W. F., & Panter, A. T. (1993). Shared meaning and the convergence among observers' personality descriptions. *Journal of Personality*, *61*, 553–585. https://doi.org/10.1111/j.1467-6494.1993.tb00782.x

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*, 1092–1122. https://doi.org/10.1037/a0021212

Cottrell, C. A., Neuberg, S. L., & Li, N. P. (2007). What do people desire in others? A sociofunctional perspective on the importance of different valued characteristics. *Journal of Personality and Social Psychology*, *92*, 208–231. https://doi.org/10.1037/0022-3514.92.2.208

Craik, K. H. (2009). *Reputation: A network interpretation*. New York, NY: Oxford University Press.

Cronbach, L. J. (1955). Processes affecting scores on understanding of others and assumed similarity. *Psychological Bulletin*, *52*, 177–193. https://doi.org/10.1037/h0044919

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126. http://dx.doi.org/10.1037/0022-3514.44.1.113

Denissen, J. J. a., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, *42*, 1285–1302. https://doi.org/10.1016/j.jrp.2008.04.002

DesJardins, N. M. L., Srivastava, S., Küfner, A. C. P., & Back, M. D. (2015). Who Attains Status? Similarities and Differences Across Social Contexts. *Social Psychological and Personality Science*, *6*, 692–700. https://doi.org/10.1177/1948550615580171

DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*, *56*, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, *8*, 100–110. https://doi.org/10.1037/1089-2680.8.2.100

Dunning, D. (1999). A Newer Look: Motivated Social Cognition and the Schematic Representation of Social Concepts. *Psychological Inquiry*, *10*, 1–11. https://doi.org/10.1207/s15327965pli1001_1

Farley, S. D. (2011). Is gossip power? The inverse relationships between gossip, power, and likability. *European Journal of Social Psychology*, *41*, 574–579. https://doi.org/10.1002/ejsp.821

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and Ostracism Promote Cooperation in Groups. *Psychological Science*, *25*, 656–664. https://doi.org/10.1177/0956797613510184

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652–670. https://doi.org/10.1037/0033-295X.102.4.652

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology, 69*, 656–672. https://doi.org/10.1037/0022-3514.69.4.656

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology, 82*, 379–398. https://doi.org/10.1037//0022-3514.82.3.379

Graham, L. T., & Gosling, S. D. (2012). Impressions of World of Warcraft players' personalities based on their usernames: Interobserver consensus but no accuracy. *Journal of Research in Personality, 46*, 599–603. https://doi.org/10.1016/j.jrp.2012.05.002

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics 3: Speech acts* (pp. 41–58). New York: Academic Press. https://doi.org/10.1111/j.1365-2664.2006.01229.x

Harari, G. M., Graham, L. T., & Gosling, S. D. (2015). Personality impressions of world of Warcraft players based on their avatars and usernames: Consensus but no accuracy. *International Journal of Gaming and Computer-Mediated Simulations, 7*, 58–73. https://doi.org/10.4018/IJGCMS.2015010104

Hogan, R. (1996). A Socioanalytic perspective on five factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical Perspectives* (pp. 163–179). New York: Guilford Press.

Human, L. J., & Biesanz, J. C. (2011). Target adjustment and self-other agreement: Utilizing trait observability to disentangle judgeability and self-knowledge. *Journal of Personality and Social Psychology, 101*, 202–216. https://doi.org/10.1037/a0023782

Human, L. J., & Biesanz, J. C. (2012). Accuracy and assumed similarity in first impressions of personality: Differing associations at different levels of analysis. *Journal of Research in Personality, 46*(1), 106–110. https://doi.org/10.1016/j.jrp.2011.10.002

Human, L. J., & Biesanz, J. C. (2013). Targeting the Good Target. *Personality and Social Psychology Review, 17*, 248–272. https://doi.org/10.1177/1088868313495593

John, O. P., & Robins, R. W. (1993). Determinants of Interjudge Agreement on Personality Traits: The Big Five Domains, Observability, Evaluativeness, and the Unique Perspective of the Self. *Journal of Personality, 61*, 521–551. https://doi.org/10.1111/j.1467-6494.1993.tb00781.x

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206–219. https://doi.org/10.1037/0022-3514.66.1.206

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*, 542–552. https://doi.org/10.1037//0021-9010.87.4.765

Kadushin, C. (2012). *Understanding social networks: Theories, concepts, and findings.* New York, NY: Oxford University Press.

Kenny, D. A. (1994). *Interpersonal Perception: A social relations analysis.* New York, NY: The Guilford Press.

Kenny, D. A. (2004). PERSON: a general model of interpersonal perception. *Personality and Social Psychology Review*, *8*, 265–280. https://doi.org/10.1207/s15327957pspr0803_3

Kwan, V. S. Y., John, O. P., Kenny, D. a, Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: an interpersonal approach. *Psychological Review*, *111*, 94–110. https://doi.org/10.1037/0033-295X.111.1.94

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, *98*, 668–682. https://doi.org/10.1037/a0018771

Leising, D., Locke, K. D., Kurzius, E., & Zimmermann, J. (2016). Quantifying the Association of Self-Enhancement Bias with Self-Ratings of Personality and Life Satisfaction. *Assessment*, *23*, 588–602. https://doi.org/10.1177/1073191115590852

MacCoun, R., & Perlmutter, S. (2015). Hide results to seek the truth. *Nature*, *526*, 187–189. https://doi.org/10.1038/526187a

Malloy, T. E., Agatstein, F., Yarlas, a, & Albright, L. (1997). Effects of communication, information overlap, and behavioral consistency on consensus in social perception. *Journal of Personality and Social Psychology*, *73*, 270–280. https://doi.org/10.1037//0022-3514.73.2.270

McAbee, S. T., & Connelly, B. S. (2016). A Multi-Rater Framework for Studying Personality: The Trait-Reputation-Identity Model. *Psychological Review*, *123*. https://doi.org/10.1037/rev0000035

Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996). The benefits of positive illusions: Idealization and the construction of satisfaction in close relationships. *Journal of Personality and Social Psychology*, *70*(1), 79–98. https://doi.org/10.1037/0022-3514.70.1.79

Muthén, L. K. & Muthén, B. O. (1998-2012). Mplus user's guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Olsen, J. A., & Kenny, D. A. (2006). Structural equation modeling with interchangeable dyads. *Psychological Methods*, *11*, 127–141. https://doi.org/10.1037/1082-989X.11.2.127

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*6, 1025-1060. https://doi.org/10.1111/1467-6494.00041

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Robins, R., & John, O. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In R. Hogan, J. Johnson, S. Briggs (Eds.), *Handbook of Personality Psychology*, 649–679. https://doi.org/10.1016/B978-012134645-4/50026-3

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/.

Saucier, G., & Goldberg, L. R. (1996). The Language of Personality: Lexical Perspectives on the Five-Factor Model. In *The five-factor model of personality: Theoretical perspectives* (pp. 21–50). New York: Guilford Press.

Saucier, G. & Srivastava, S. (2015). What makes a good structural model of personality? Evaluating the big five and alternatives. In M. Mikulincer, P. R. Shaver, M. L. Cooper & R. J. Larsen (Eds.), *Handbook of personality and social psychology. Vol. 3: Personality processes and individual differences* (pp. 283-305). Washington, DC: APA Books.

Soto, C. J., & John, O. P. (2017). The next big five inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*. 117-143. https://doi.org/http://dx.doi.org/10.1037/pspp0000096

Srivastava, S. (2010). The Five-Factor Model Describes the Structure of Social Perceptions. *Psychological Inquiry*, *21*, 69–75. https://doi.org/10.1080/10478401003648815

Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of Brief to Medium-Length Big Five and Big Six Personality Questionnaires. *Psychological Assessment*, *23*(4), 995–1009. https://doi.org/10.1037/a0024165

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*, 281–300. https://doi.org/10.1037/a0017908

Wood, D. (2015). Testing the lexical hypothesis: Are socially important traits more densely reflected in the English lexicon? *Journal of Personality and Social Psychology*, *108*(2), 317–335. https://doi.org/10.1037/a0038343

Wood, D., Tov, W., Costello, C., Tov, W., & Costello, C. (2015). What a _____ Thing to Do ! Formally Characterizing Actions by Their Expected Effects. *Journal of Personality and Social Psychology*, *108*, 953-976. http://dx.doi.org/10.1037/pspp0000030

Wortman, J., & Wood, D. (2011). The personality traits of liked people. *Journal of Research in Personality*, *45*, 519–528. https://doi.org/10.1016/j.jrp.2011.06.006

Wu, J., Balliet, D., & Van Lange, P. A. M. (2015). When Does Gossip Promote Generosity? Indirect Reciprocity Under the Shadow of the Future. *Social Psychological and Personality Science*, *6*(8), 923–930. https://doi.org/10.1177/1948550615595272

Table S1

Coefficient Alphas for Big Six Ratings

Report Type	Domain	Study 1 α	Study 2 α
Self	agreeableness	.76	.76
	conscientiousness	.83	.86
	honesty-propriety	.61	.62
	neuroticism	.89	.88
	extraversion	.87	.87
	openness	.81	.81
P1	agreeableness	.80	.82
	conscientiousness	.83	.88
	honesty-propriety	.66	.76
	neuroticism	.75	.84
	extraversion	.88	.88
	openness	.78	.87
P2	agreeableness	.84	.83
	conscientiousness	.89	.91
	honesty-propriety	.75	.79
	neuroticism	.87	.83
	extraversion	.89	.91
	openness	.83	.77

Note. For self-, P1-, and P2-reports, each of the Big Five was measured with 12 items and honesty-propriety with 8 items.

Table S2

Example Variance-Covariance Matrix

	B [C]	D [A]	A [A]	C [C]	A [C]	C [A]
B [C]	P2					
D [A]	h	P2				
A [A]	j	e	self			
C [C]	e	j	i	self		
A [C]	f	m	l	g	P1	
C [A]	m	f	g	l	k	P1

Note. Letters in the row and column labels correspond to the names in the graphical depiction of the study design in Figure 1a. The perceiver's procedural role appears outside of the brackets and the target's procedural role appears inside of the brackets, e.g., A [C] is A's perception of C. Cells with the same letter are conceptually identical and pooled in our analyses (this includes variances with the same label). Intercept equality constraints are not pictured, but ratings which have equality constraints on variances would also have equality constraints on intercepts. Cells with gray backgrounds are unique (i.e., they do not have a conceptually identical estimate elsewhere in the model).

Table S3

Pooled Means and Standard Deviations among Perceptions in Study 1

	T(T)	*P1(T)*	*P2(T)*
	M (SD)	*M (SD)*	*M (SD)*
A	3.84 (0.58)	3.96 (0.51)	3.66 (0.52)
C	3.49 (0.65)	3.69 (0.55)	3.54 (0.61)
HP	3.70 (0.58)	3.81 (0.49)	3.42 (0.54)
N	2.92 (0.85)	2.55 (0.49)	2.76 (0.60)
E	3.43 (0.73)	3.19 (0.72)	3.23 (0.74)
O	3.81 (0.59)	3.46 (0.47)	3.31 (0.48)

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. T(T) = the target's self-report. P1(T) = P1's rating of the target. P2(T) = P2's rating of the target.

Table S4

Extraversion Moderator Analyses on Hearsay Consensus

	Target's extraversion			P1's extraversion			P2's extraversion		
	β	95% CI	*p*	*β*	95% CI	*p*	*β*	95% CI	*p*
A	.19	[.00, .38]	.051	-.03	[-.22, .16]	.725	-.09	[-.28, .11]	.380
C	.00	[-.18, .18]	.981	.17	[-.01, .34]	.058	-.11	[-.28, .07]	.224
HP	.15	[-.03, .32]	.096	-.08	[-.26, .10]	.391	-.14	[-.32, .04]	.122
N	.04	[-.14, .21]	.675	.11	[-.06, .28]	.218	.07	[-.10, .25]	.397
E	-.07	[-.24, .09]	.364	.12	[-.04, .27]	.131	.00	[-.15, .15]	.975
O	.17	[.00, .35]	.053	.06	[-.12, .24]	.534	.02	[-.17, .21]	.828

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. The column labelled *β* contains the standardized regression slope for the interaction term from the path analyses. The column labelled 'Target's extraversion' contains results from analyses of the target's level of extraversion as a moderator of hearsay consensus. The column labelled 'P1's extraversion' contains results from analyses of P1's level of extraversion as a moderator of hearsay consensus. The column labelled 'P2's extraversion' contains results from analyses of P2's level of extraversion as a moderator of hearsay consensus.

Table S5

Extraversion Moderator Analyses on Hearsay Accuracy

	Target's extraversion			P1's extraversion			P2's extraversion		
	β	95% CI	*p*	*β*	95% CI	*p*	*β*	95% CI	*p*
A	-.05	[-.24, .14]	.612	.00	[-.20, .19]	.987	-.11	[-.30, .07]	.236
C	.03	[-.16, .22]	.746	.16	[-.03, .36]	.106	-.08	[-.27, .11]	.423
HP	.15	[-.05, .34]	.136	.04	[-.17, .25]	.719	-.05	[-.25, .16]	.668
N	.10	[-.06, .27]	.225	.16	[-.02, .34]	.079	.06	[-.13, .24]	.565
E	--	--	--	-.01	[-.20, .17]	.878	-.03	[-.20, .15]	.778
O	.19	[.02, .37]	.032	.17	[-.02, .37]	.080	-.01	[-.20, .19]	.951

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. The column labelled *β* contains the standardized regression slope for the interaction term from the path analyses. The column labelled 'Target's extraversion' contains results from analyses of the target's level of extraversion as a moderator of hearsay accuracy. The column labelled 'P1's extraversion' contains results from analyses of P1's level of extraversion as a moderator of hearsay accuracy. The column labelled 'P2's extraversion' contains results from analyses of P2's level of extraversion as a moderator of hearsay accuracy.

Table S6

P2 Empathy Moderator Analyses

		Hearsay Consensus			Hearsay Accuracy		
Empathy Subscale		*β*	95% CI	*p*	*β*	95% CI	*p*
Perspective Taking	A	.19	[.01, .37]	.044	.05	[-.15, .24]	.644
	C	.11	[-.07, .30]	.229	.18	[-.02, .38]	.070
	HP	.07	[-.11, .25]	.454	.17	[-.03, .37]	.093
	N	-.05	[-.24, .14]	.581	.10	[-.08, .29]	.286
	E	.02	[-.16, .19]	.854	.12	[-.06, .31]	.200
	O	.19	[.01, .37]	.044	.14	[-.05, .34]	.153
Empathic Concern	A	-.16	[-.35, .04]	.110	-.03	[-.23, .17]	.758
	C	-.11	[-.28, .07]	.244	.02	[-.19, .24]	.824
	HP	-.09	[-.27, .09]	.344	-.20	[-.41, .01]	.062
	N	-.11	[-.29, .08]	.252	-.04	[-.22, .15]	.711
	E	.04	[-.13, .20]	.672	.04	[-.13, .22]	.636
	O	.10	[-.09, .29]	.304	.00	[-.21, .20]	.968
Personal Distress	A	-.06	[-.27, .15]	.592	.03	[-.18, .24]	.757
	C	-.02	[-.20, .15]	.794	.10	[-.11, .31]	.346
	HP	.00	[-.18, .19]	.971	-.11	[-.32, .10]	.286
	N	.15	[-.03, .34]	.110	.04	[-.16, .24]	.684
	E	.05	[-.12, .23]	.549	-.04	[-.23, .15]	.683
	O	.09	[-.11, .28]	.380	-.06	[-.27, .16]	.611
Fantasy	A	-.07	[-.27, .14]	.516	-.08	[-.29, .13]	.447
	C	-.05	[-.23, .13]	.582	-.15	[-.35, .05]	.129
	HP	-.10	[-.27, .08]	.271	-.10	[-.32, .12]	.356
	N	-.10	[-.28, .08]	.261	-.03	[-.23, .17]	.751
	E	-.03	[-.20, .13]	.698	-.05	[-.24, .13]	.567
	O	.11	[-.08, .30]	.273	.12	[-.10, .35]	.282

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. The column labelled *β* contains the standardized regression slope for the interaction term from the path analyses. The columns under the 'Hearsay Consensus' header show the results assessing P2 empathy as a moderator for hearsay consensus; the columns under the 'Hearsay Accuracy' heading show results for P2 empathy as a moderator for hearsay accuracy.

Table S7

Model Comparisons for Testing Condition Differences for Extraversion

Equality Constraint	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*
Hearsay Consensus	-0.65	-3.29	1.35	.244
Hearsay Accuracy	-1.88	-4.52	0.12	.727
Direct Accuracy	0.22	-2.42	2.22	.136

Note. Δ_{AIC} and Δ_{BIC} are the change in AIC and BIC (respectively) resulting from the constraint; negative values indicate that the constraint leads to better fit, and positive values indicate that the constraint leads to worse fit. χ_d^2 is the difference in χ^2 between the model with and without constraints. Each model is compared to the baseline model (i.e., the model with no cross-condition equality constraints). Equality constraints were imposed across the Status and Control conditions for Extraversion. The difference in df_{model} for each comparison is 1.

Table S8

Model Comparisons for Testing Condition Differences for Agreeableness

Equality Constraint	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*
Hearsay Consensus	-1.89	-4.54	0.11	.741
Hearsay Accuracy	-1.06	-3.71	0.94	.333
Direct Accuracy	-1.99	-4.64	0.01	.930

Note. Δ_{AIC} and Δ_{BIC} are the change in AIC and BIC (respectively) resulting from the constraint; negative values indicate that the constraint leads to better fit, and positive values indicate that the constraint leads to worse fit. χ_d^2 is the difference in χ^2 between the model with and without constraints. Each model is compared to the baseline model (i.e., the model with no cross-condition equality constraints). Equality constraints were imposed across the Affiliation and Control conditions for Agreeableness. The difference in df_{model} for each comparison is 1.

Table S9

Model Comparisons Testing for Between Study Differences in Full Variance-Covariance and Mean Structure

	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*
agreeableness	-15.82	-63.47	14.18	.512
conscientiousness	-5.04	-52.68	24.96	.050
honesty-propriety	-11.61	-59.25	18.39	.243
neuroticism	-8.08	-55.72	21.92	.110
extraversion	-20.50	-68.14	9.50	.850
openness	-14.31	-61.95	15.69	.403

Note. The df_{model} is 24 for the model allowing studies to differ and 39 for the model pooling across studies; the *df* for the model comparison is therefore 15. Δ_{AIC} and Δ_{BIC} = the difference in AIC and BIC (respectively) between the model allowing studies to differ and the model pooling across studies; negative values indicate better fit for the pooled (integrated) model.

Table S10

Pooled Means and Standard Deviations among Perceptions in Study 2

	Integrated			*Study 2*		
	T(T)	*P1(T)*	*P2(T)*	*T(T)*	*P1(T)*	*P2(T)*
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
A	3.86 (0.55)	3.91 (0.54)	3.71 (0.54)	3.88 (0.54)	3.88 (0.56)	3.74 (0.54)
C	3.52 (0.63)	3.70 (0.60)	3.66 (0.64)	3.55 (0.62)	3.70 (0.64)	3.75 (0.65)
HP	3.70 (0.57)	3.76 (0.55)	3.50 (0.56)	3.69 (0.56)	3.72 (0.58)	3.56 (0.56)
N	2.90 (0.81)	2.57 (0.56)	2.71 (0.56)	2.88 (0.78)	2.59 (0.59)	2.68 (0.53)
E	3.39 (0.73)	3.21 (0.71)	3.25 (0.78)	3.37 (0.74)	3.23 (0.71)	3.26 (0.81)
O	3.80 (0.56)	3.44 (0.53)	3.33 (0.45)	3.80 (0.55)	3.44 (0.57)	3.35 (0.43)

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. T(T) = the target's self-report. P1(T) = P1's rating of the target. P2(T) = P2's rating of the target.

Table S11

Model Comparisons Testing for Between Study Differences for Distinctive Hearsay Accuracy

	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*
Agreeableness	-7.55	-36.14	10.45	.316
Conscientiousness	-3.21	-31.79	14.79	.097
Honesty-Propriety	-2.84	-31.43	15.16	.087
Neuroticism	-3.84	-32.42	14.16	.117
Extraversion	-14.18	-42.77	3.82	.923
Openness	-4.67	-33.26	13.33	.148

Note. The df_{model} is 34 for the model allowing studies to differ and 43 for the model pooling across studies; the *df* for the model comparison is therefore 9. Δ_{AIC} and Δ_{BIC} = the difference in AIC and BIC (respectively) between the model allowing studies to differ and the model pooling across studies; negative values indicate better fit for the pooled (integrated) model.

Table S12

Model Comparisons Testing for Between Study Differences for Target, P1, and P2 Extraversion Moderating Hearsay Consensus

	Target Extraversion				P1 Extraversion				P2 Extraversion			
	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*
Agreeableness	-10.97	-45.90	11.03	.440	-13.60	-48.54	8.40	.678	-10.38	-45.32	11.62	.393
Conscientiousness	-0.67	-35.61	21.33	.030	-7.71	-42.65	14.29	.217	-3.03	-37.97	18.97	.062
Honesty-Propriety	-0.84	-35.78	21.16	.032	-3.97	-38.91	18.03	.081	-5.50	-40.43	16.50	.123
Neuroticism	-6.39	-41.33	15.61	.156	-4.12	-39.06	17.88	.084	-2.30	-37.24	19.70	.050
Extraversion	-17.64	-52.57	4.36	.958	-15.77	-50.71	6.23	.857	-12.49	-47.43	9.51	.575
Openness	-10.91	-45.85	11.09	.436	-10.39	-45.33	11.61	.394	6.74	-28.19	28.74	.002

Note. The df_{model} is 64 for the model allowing studies to differ and 75 for the model pooling across studies; the *df* for the model comparison is therefore 11. Δ_{AIC} and Δ_{BIC} = the difference in AIC and BIC (respectively) between the model allowing studies to differ and the model pooling across studies; negative values indicate better fit for the pooled (integrated) model.

Table S13

Model Comparisons Testing for Between Study Differences for Target, P1, and P2 Extraversion Moderating Hearsay Accuracy

	Target Extraversion				P1 Extraversion				P2 Extraversion			
	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*
Agreeableness	-8.68	-43.62	13.32	.273	-10.52	-45.46	11.48	.404	-12.77	-47.71	9.23	.601
Conscientiousness	-2.46	-37.40	19.54	.052	-10.69	-45.63	11.31	.418	-9.08	-44.01	12.92	.298
Honesty-Propriety	-8.17	-43.10	13.83	.242	-11.53	-46.47	10.47	.489	-14.39	-49.33	7.61	.748
Neuroticism	-13.13	-48.07	8.87	.634	-9.62	-44.56	12.38	.336	-10.62	-45.56	11.38	.412
Extraversion	--		--	--	-9.86	-32.10	4.14	.764	-13.82	-48.76	8.18	.697
Openness	-14.26	-49.20	7.74	.737	-16.24	-51.18	5.76	.889	5.51	-29.43	27.51	.004

Note. The df_{model} is 64 for the model allowing studies to differ and 75 for the model pooling across studies; the *df* for the model comparison is therefore 11. The one exception for this is P1 Extraversion moderating hearsay accuracy for Extraversion; since P1s' self-reported Extraversion is an outcome in one equation and a moderator in the other, the df_{model} is 40 in the model allowing studies to differ and 47 for the model pooling across studies, resulting in a *df* of 7 for the comparison. Δ_{AIC} and Δ_{BIC} = the difference in AIC and BIC (respectively) between the model allowing studies to differ and the model pooling across studies; negative values indicate better AIC or BIC values for the pooled model.

Table S14

Extraversion Moderator Analyses on Hearsay Consensus in Integrated Data

	Target's extraversion			P1's extraversion			P2's extraversion		
	β	95% CI	*p*	*β*	95% CI	*p*	*β*	95% CI	*p*
A	.16	[.05, .28]	.005	-.02	[-.13, .10]	.775	-.13	[-.24, -.01]	.031
C	.03	[-.08, .13]	.641	.08	[-.02, .19]	.117	-.12	[-.23, -.01]	.031
HP	.06	[-.05, .17]	.302	-.03	[-.14, .09]	.671	-.08	[-.20, .04]	.191
N	.05	[-.05, .16]	.333	.07	[-.04, .17]	.226	.07	[-.04, .17]	.230
E	.00	[-.10, .09]	.942	.10	[.01, .20]	.027	-.03	[-.13, .06]	.476
O	.09	[-.02, .20]	.120	-.02	[-.14, .09]	.682	-.04	[-.16, .07]	.483

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. The column labelled *β* contains the standardized regression slope for the interaction term from the path analyses. The column labelled 'Target's extraversion' contains results from analyses of the target's level of extraversion as a moderator of hearsay consensus. The column labelled 'P1's extraversion' contains results from analyses of P1's level of extraversion as a moderator of hearsay consensus. The column labelled 'P2's extraversion' contains results from analyses of P2's level of extraversion as a moderator of hearsay consensus.

Table S15

Extraversion Moderator Analyses on Hearsay Consensus in Study 2 Data

	Target's extraversion			P1's extraversion			P2's extraversion		
	β	95% CI	*p*	*β*	95% CI	*p*	*β*	95% CI	*p*
A	.16	[.02, .30]	.022	-.01	[-.15, .13]	.899	-.15	[-.30, .00]	.044
C	.03	[-.11, .17]	.688	.06	[-.08, .20]	.386	-.17	[-.31, -.03]	.020
HP	.03	[-.12, .18]	.683	-.01	[-.16, .14]	.885	-.03	[-.19, .13]	.677
N	.07	[-.06, .21]	.302	.02	[-.12, .15]	.790	.07	[-.07, .21]	.333
E	.03	[-.09, .14]	.663	.09	[-.02, .21]	.107	-.04	[-.16, .07]	.484
O	.05	[-.10, .20]	.499	-.06	[-.21, .08]	.394	-.09	[-.24, .06]	.245

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. The column labelled *β* contains the standardized regression slope for the interaction term from the path analyses. The column labelled 'Target's extraversion' contains results from analyses of the target's level of extraversion as a moderator of hearsay consensus. The column labelled 'P1's extraversion' contains results from analyses of P1's level of extraversion as a moderator of hearsay consensus. The column labelled 'P2's extraversion' contains results from analyses of P2's level of extraversion as a moderator of hearsay consensus.

Table S16

Extraversion Moderator Analyses on Hearsay Accuracy in Integrated Data

	Target's extraversion			P1's extraversion			P2's extraversion		
	β	95% CI	*p*	*β*	95% CI	*p*	*β*	95% CI	*p*
A	-.03	[-.15, .09]	.648	.08	[-.05, .20]	.223	-.05	[-.18, .07]	.386
C	-.03	[-.14, .09]	.632	.10	[-.02, .22]	.107	-.11	[-.23, .01]	.072
HP	.05	[-.08, .17]	.439	-.04	[-.16, .09]	.565	-.08	[-.20, .05]	.239
N	.07	[-.03, .17]	.188	.08	[-.04, .19]	.188	.06	[-.06, .18]	.313
E	--	--	--	.19	[.16, .23]	<.001	-.13	[-.25, -.02]	.019
O	.07	[-.04, .19]	.203	.10	[-.02, .22]	.103	-.09	[-.21, .03]	.153

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. The column labelled *β* contains the standardized regression slope for the interaction term from the path analyses. The column labelled 'Target's extraversion' contains results from analyses of the target's level of extraversion as a moderator of hearsay accuracy. The column labelled 'P1's extraversion' contains results from analyses of P1's level of extraversion as a moderator of hearsay accuracy. The column labelled 'P2's extraversion' contains results from analyses of P2's level of extraversion as a moderator of hearsay accuracy.

Table S17

Extraversion Moderator Analyses on Hearsay Accuracy in Study 2 Data

	Target's extraversion			P1's extraversion			P2's extraversion		
	β	95% CI	*p*	*β*	95% CI	*p*	*β*	95% CI	*p*
A	-.05	[-.20, .10]	.523	.15	[.00, .30]	.050	-.04	[-.21, .13]	.616
C	-.03	[-.18, .11]	.657	.07	[-.08, .23]	.347	-.15	[-.30, .01]	.064
HP	-.03	[-.19, .14]	.751	-.07	[-.23, .09]	.407	-.10	[-.27, .06]	.220
N	.05	[-.08, .19]	.430	.00	[-.15, .15]	1.000	.08	[-.08, .23]	.325
E	--	--	--	.20	[.15, .25]	<.001	-.21	[-.35, -.08]	.002
O	.01	[-.14, .16]	.916	.05	[-.10, .20]	.479	-.16	[-.32, -.01]	.032

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. The column labelled *β* contains the standardized regression slope for the interaction term from the path analyses. The column labelled 'Target's extraversion' contains results from analyses of the target's level of extraversion as a moderator of hearsay accuracy. The column labelled 'P1's extraversion' contains results from analyses of P1's level of extraversion as a moderator of hearsay accuracy. The column labelled 'P2's extraversion' contains results from analyses of P2's level of extraversion as a moderator of hearsay accuracy.

Table S18

Model Comparisons Testing for Between Study Differences for P2 Empathy Moderator Analyses

		Hearsay Consensus				Hearsay Accuracy			
Empathy	Domain	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*	Δ_{AIC}	Δ_{BIC}	χ_d^2	*p*
Perspective Taking	A	3.66	-31.52	25.66	.007	-14.85	-50.03	7.15	.787
	C	-7.81	-43.00	14.19	.223	-9.82	-45.01	12.18	.350
	HP	2.94	-32.25	24.94	.009	-11.73	-46.91	10.27	.506
	N	-7.11	-42.29	14.89	.188	-13.94	-49.12	8.06	.708
	E	-17.26	-52.44	4.74	.943	-14.74	-49.93	7.26	.778
	O	-3.21	-38.39	18.79	.065	-13.52	-48.71	8.48	.670
Empathic Concern	A	-6.70	-41.88	15.30	.169	-15.14	-50.33	6.86	.811
	C	-9.32	-44.50	12.68	.314	-8.84	-44.03	13.16	.283
	HP	-2.24	-37.43	19.76	.049	-7.00	-42.18	15.00	.182
	N	-4.11	-39.29	17.89	.084	-10.39	-45.58	11.61	.394
	E	-15.92	-51.11	6.08	.868	-8.27	-43.45	13.73	.248
	O	3.44	-31.74	25.44	.008	-6.32	-41.50	15.68	.153
Personal Distress	A	-6.16	-41.35	15.84	.147	-13.92	-49.10	8.08	.706
	C	-3.87	-39.06	18.13	.079	-9.68	-44.86	12.32	.340
	HP	-0.39	-35.58	21.61	.028	-10.87	-46.06	11.13	.433
	N	5.27	-29.91	27.27	.004	-10.66	-45.84	11.34	.415
	E	-9.77	-44.95	12.23	.347	-12.52	-47.70	9.48	.577
	O	-3.19	-38.38	18.81	.065	-9.53	-44.71	12.47	.329
Fantasy	A	-8.33	-43.52	13.67	.252	-9.90	-45.09	12.10	.356
	C	-7.60	-42.78	14.40	.212	-8.29	-43.48	13.71	.250
	HP	-4.77	-39.96	17.23	.101	-10.90	-46.09	11.10	.435
	N	-5.52	-40.71	16.48	.124	-13.57	-48.76	8.43	.675
	E	-12.11	-47.29	9.89	.540	-11.05	-46.23	10.95	.448
	O	-0.02	-35.20	21.98	.025	-14.97	-50.16	7.03	.797

Note. A = Agreeableness; C = Conscientiousness; HP = Honesty-Propriety; N = Neuroticism; E = Extraversion; O = Openness. The df_{model} is 64 for the model allowing studies to differ and 75 for the model pooling across studies; the *df* for the model comparison is therefore 11. $\Delta_{AIC \text{ and}}$ Δ_{BIC} = the difference in AIC and BIC (respectively) between the model allowing studies to differ and the model pooling across studies; negative values indicate better fit for the pooled (integrated) model.

Table S19

P2 Empathy Moderator Analyses in Integrated Data

		Hearsay Consensus			Hearsay Accuracy		
Empathy Subscale		*β*	95% CI	*p*	*β*	95% CI	*p*
Perspective Taking	A	.09	[-.03, .21]	.140	.02	[-.10, .15]	.722
	C	.01	[-.10, .12]	.854	.08	[-.05, .20]	.227
	HP	.00	[-.12, .12]	.968	.17	[.05, .30]	.007
	N	.02	[-.09, .13]	.706	.12	[.00, .23]	.053
	E	-.02	[-.12, .08]	.659	-.02	[-.14, .10]	.718
	O	.04	[-.07, .16]	.462	.05	[-.08, .18]	.443
Empathic Concern	A	-.14	[-.26, -.02]	.021	.06	[-.06, .19]	.320
	C	-.15	[-.25, -.04]	.007	-.04	[-.16, .09]	.576
	HP	-.15	[-.26, -.03]	.015	.05	[-.08, .18]	.454
	N	-.11	[-.21, .00]	.055	.03	[-.09, .15]	.586
	E	-.03	[-.13, .07]	.567	-.10	[-.22, .01]	.072
	O	-.02	[-.14, .10]	.775	-.02	[-.14, .11]	.775
Personal Distress	A	.01	[-.12, .13]	.910	.08	[-.05, .21]	.220
	C	-.09	[-.20, .02]	.119	.06	[-.06, .19]	.327
	HP	-.06	[-.18, .06]	.290	-.01	[-.14, .12]	.862
	N	.04	[-.07, .15]	.495	.00	[-.12, .13]	.976
	E	.09	[-.01, .19]	.084	-.01	[-.13, .10]	.826
	O	.06	[-.05, .18]	.302	.05	[-.08, .17]	.467
Fantasy	A	-.08	[-.20, .04]	.191	.08	[-.04, .21]	.200
	C	-.08	[-.19, .03]	.172	-.06	[-.19, .06]	.306
	HP	-.10	[-.21, .02]	.094	.07	[-.06, .19]	.318
	N	-.10	[-.20, .01]	.082	-.01	[-.13, .11]	.867
	E	-.02	[-.12, .08]	.669	-.07	[-.18, .04]	.226
	O	-.06	[-.18, .05]	.295	.03	[-.10, .15]	.682

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. The column labelled *β* contains the standardized regression slope for the interaction term from the path analyses. The columns under the 'Hearsay Consensus' header show the results assessing P2 empathy as a moderator for hearsay consensus; the columns under the 'Hearsay Accuracy' heading show results for P2 empathy as a moderator for hearsay accuracy.

Table S20

P2 Empathy Moderator Analyses in Study 2 Data

		Hearsay Consensus			Hearsay Accuracy		
Empathy Subscale		*β*	95% CI	*p*	*β*	95% CI	*p*
Perspective Taking	A	.06	[-.10, .21]	.472	.00	[-.16, .17]	.978
	C	-.03	[-.17, .11]	.677	.02	[-.14, .18]	.798
	HP	-.03	[-.19, .13]	.747	.17	[.01, .33]	.042
	N	.03	[-.11, .18]	.639	.14	[-.02, .29]	.086
	E	-.05	[-.17, .08]	.483	-.11	[-.26, .04]	.144
	O	-.02	[-.18, .13]	.757	-.03	[-.19, .14]	.736
Empathic Concern	A	-.11	[-.26, .04]	.155	.11	[-.05, .27]	.195
	C	-.15	[-.30, -.01]	.036	-.09	[-.25, .07]	.275
	HP	-.15	[-.31, .00]	.049	.17	[.01, .33]	.043
	N	-.08	[-.23, .06]	.271	.05	[-.12, .22]	.548
	E	-.06	[-.18, .06]	.348	-.20	[-.34, -.06]	.005
	O	-.04	[-.19, .11]	.580	-.04	[-.20, .12]	.604
Personal Distress	A	.03	[-.13, .18]	.726	.12	[-.05, .28]	.165
	C	-.12	[-.26, .02]	.091	.04	[-.11, .20]	.596
	HP	-.10	[-.25, .06]	.219	.05	[-.12, .22]	.577
	N	-.01	[-.15, .13]	.906	-.02	[-.18, .14]	.817
	E	.10	[-.02, .22]	.109	-.01	[-.17, .14]	.846
	O	.02	[-.13, .17]	.799	.08	[-.08, .24]	.308
Fantasy	A	-.07	[-.23, .08]	.338	.14	[-.02, .30]	.080
	C	-.08	[-.22, .06]	.273	.00	[-.15, .16]	.973
	HP	-.10	[-.26, .05]	.193	.14	[-.03, .30]	.101
	N	-.11	[-.24, .03]	.124	.00	[-.16, .16]	.991
	E	-.02	[-.14, .11]	.792	-.09	[-.23, .06]	.254
	O	-.15	[-.30, -.01]	.034	-.02	[-.17, .13]	.791

Note. A = agreeableness; C = conscientiousness; HP = honesty-propriety; N = neuroticism; E = extraversion; O = openness. The column labelled *β* contains the standardized regression slope for the interaction term from the path analyses. The columns under the 'Hearsay Consensus' header show the results assessing P2 empathy as a moderator for hearsay consensus; the columns under the 'Hearsay Accuracy' heading show results for P2 empathy as a moderator for hearsay accuracy.

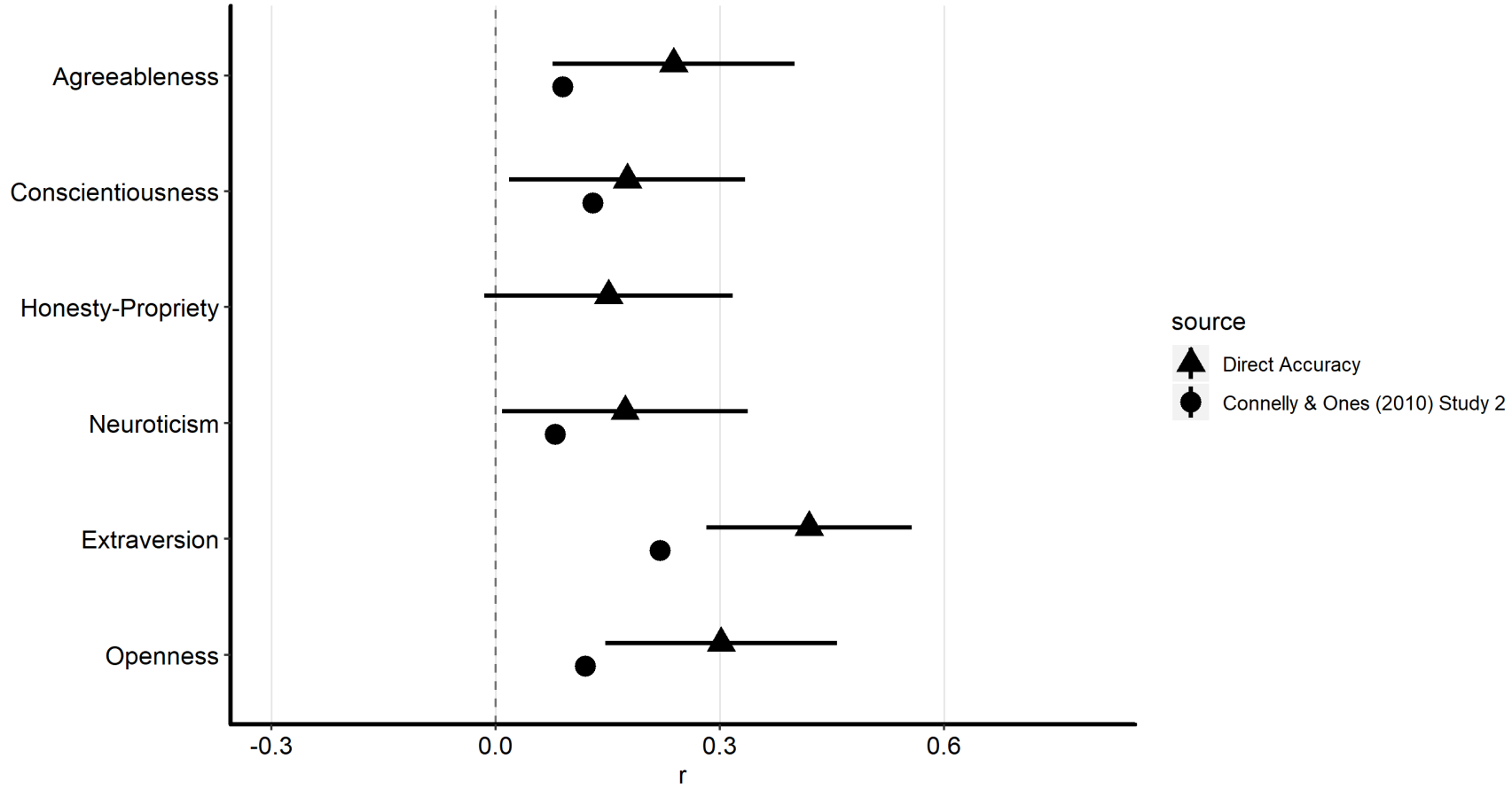
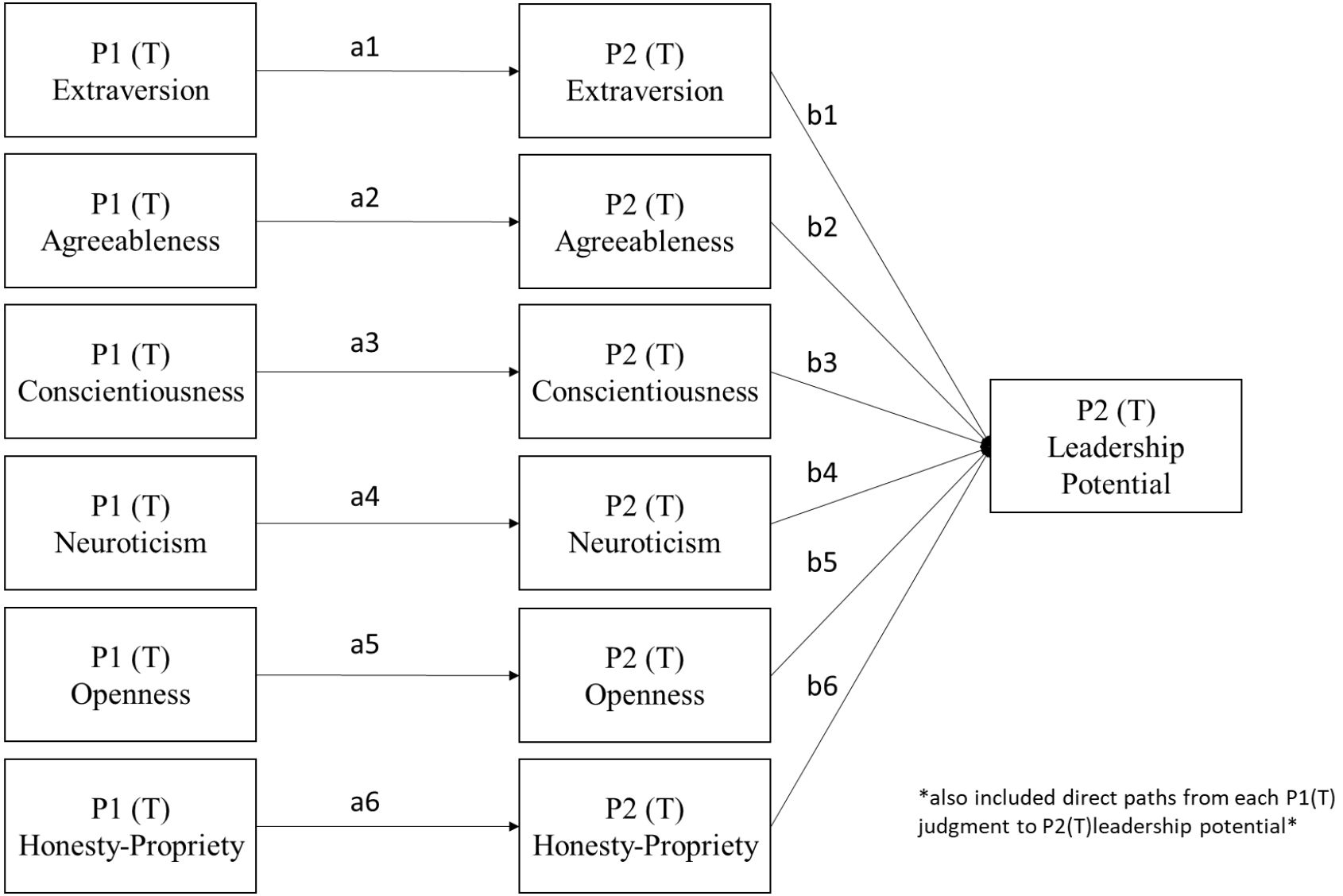


Figure S1. Direct accuracy for the Big Six in Study 1. Triangles represent point estimates for correlations, and the lines through them represent the 95% CI's around those estimates. Circles represent meta-analytic estimates of self-other agreement among previously unacquainted participants, obtained from Study 2 of Connelly & Ones's (2010) meta-analysis on consensus and self-other agreement.



Supplemental Figure S2. Path model predicting hearsay reputations of leadership potential from hearsay and direct reputations of all Big Six. The model also included direct effects from each P1(T) judgement to P2(T) leadership potential and tested each indirect effect (not depicted).

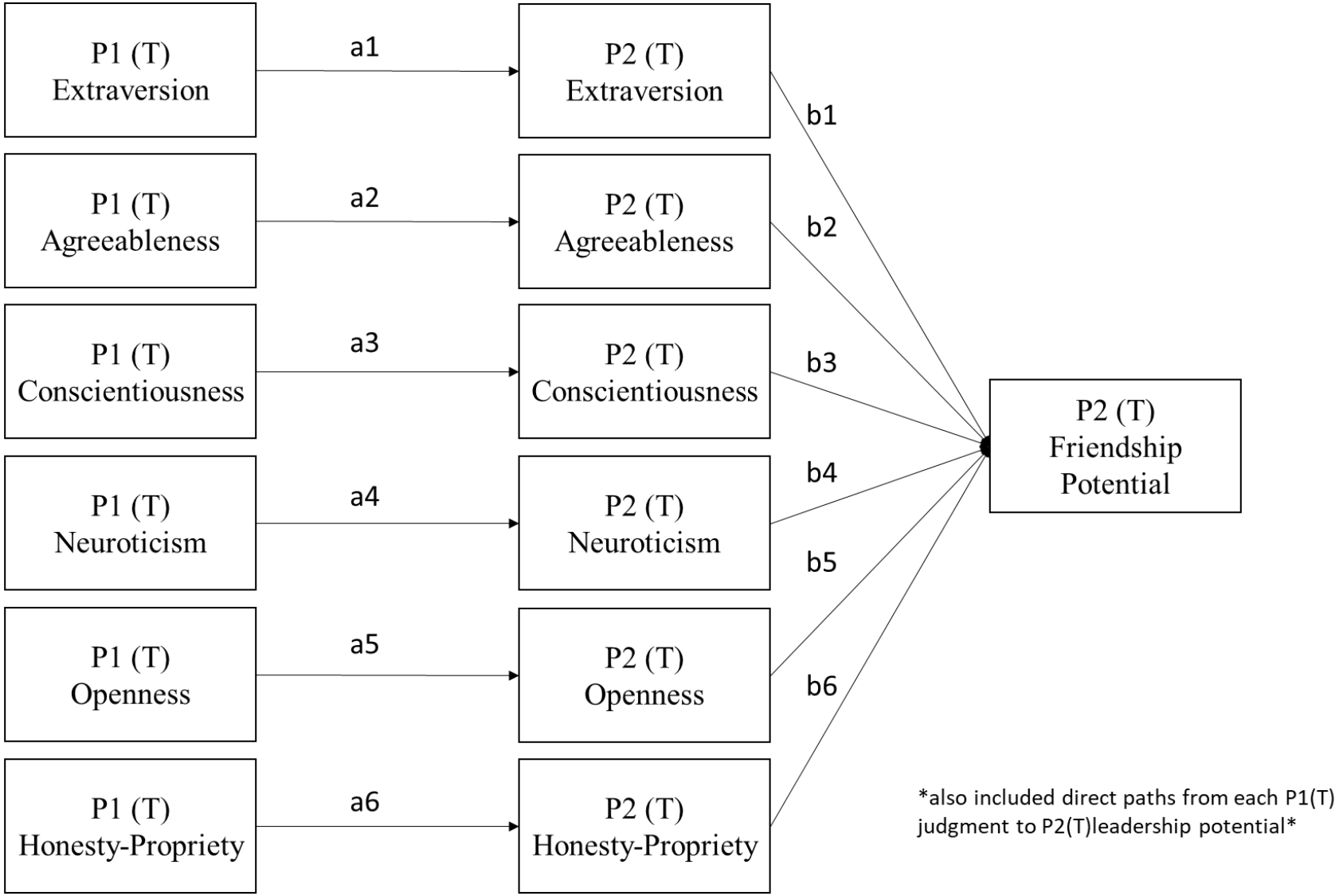


Figure S3. Path model predicting hearsay reputations of friendship potential from hearsay and direct reputations of all Big Six. The model also included direct effects from each P1(T) judgement to P2(T) leadership potential and tested each indirect effect (not depicted).

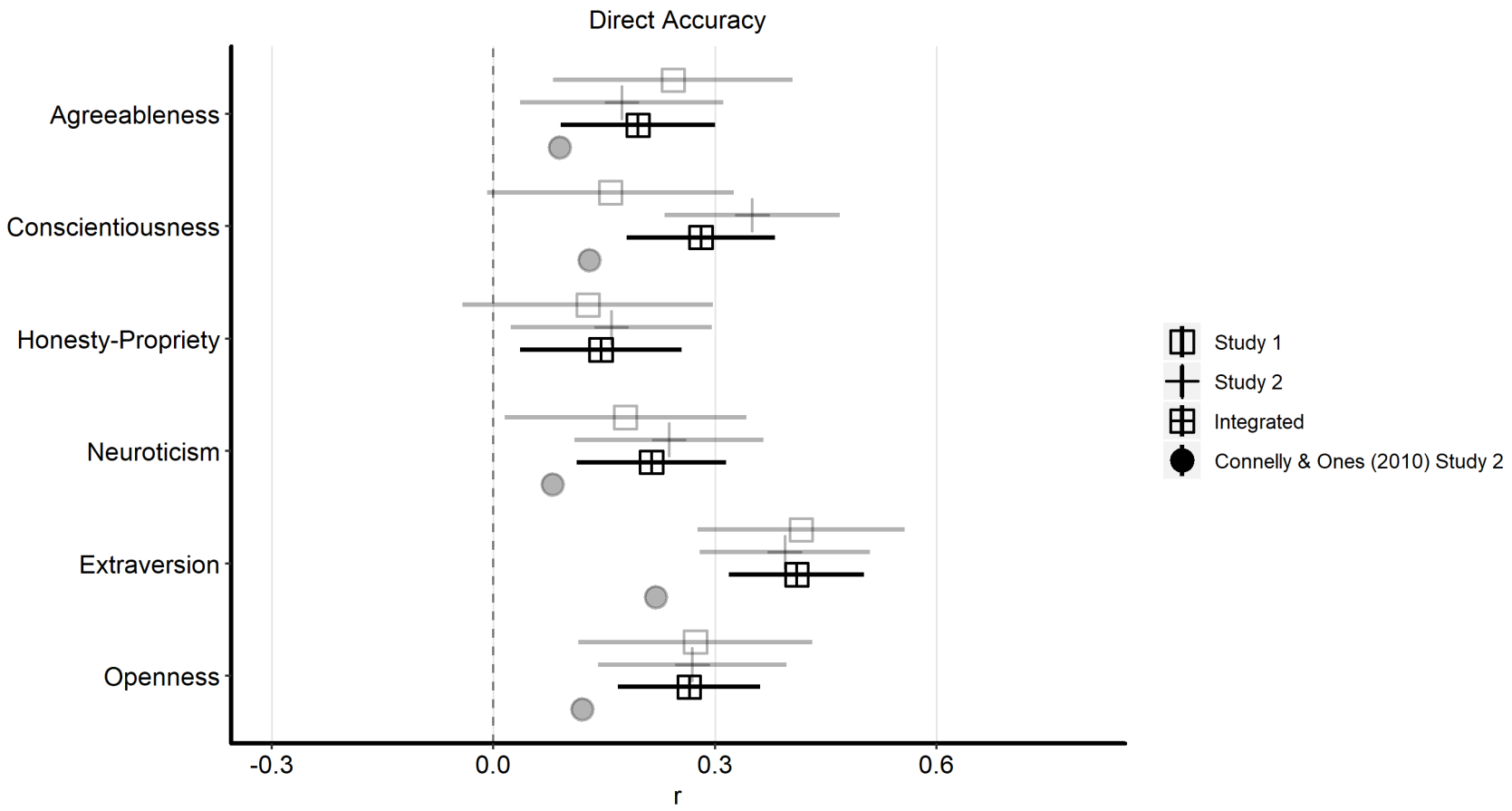


Figure S4. Study 1, Study 2, and integrated estimates of direct accuracy for the Big Six. Study 1 (box) estimates are based just on the Study 1 sample; Study 2 (cross) estimates are based just on the Study 2 sample (collapsing across condition); integrated estimates (crossed-box) are based on data from both Studies. Circles represent meta-analytic estimates of self-other agreement among previously unacquainted participants, obtained from Study 2 of Connelly & Ones's (2010) meta-analysis on consensus and self-other agreement.