Title
Intentional and Unintentional Empathy for Pain Among Physicians and Nonphysicians

Journal
Social Psychological and Personality Science, 10(4)

ISSN
1948-5506

Authors
Spring, Victoria L
Cameron, C Daryl
McKee, Stephanie
et al.

Publication Date
2019-05-01

DOI
10.1177/1948550618771890

Peer reviewed

Article

Social Psychological and Personality Science
1-9

DOI: 10.1177/1948550618771890
journals.sagepub.com/home/spp
SAGE

Intentional and Unintentional Empathy for Pain Among Physicians and Nonphysicians

Victoria L. Spring[1], C. Daryl Cameron[1], Stephanie McKee[2], and Andrew R. Todd[3]

Abstract

Empathy can be both beneficial and costly. This trade-off is pertinent for physicians who must care for patients while maintaining emotional distance to avoid burnout. Prior work using self-report and neurophysiological measures has found mixed evidence for differences in empathy between physicians and nonphysicians. We used implicit measurement and multinomial modeling to examine intentional empathy (IE) and unintentional empathy (UE) for pain among physicians and demographically matched nonphysicians. Relative to nonphysicians, physicians displayed greater ability to judge the painfulness of target experiences (i.e., IE). Contrary to some prior work, however, physicians and nonphysicians displayed comparable spontaneous resonance with distracter experiences (i.e., UE). These findings suggest that physicians may be more likely than nonphysicians to empathize with others' pain when empathy aligns with their overt goals.

Keywords

empathy, health care, implicit social cognition, multinomial modeling, pain

Feeling as others feel is an experience most of us have shared. This propensity, known as empathy, is a frequent conversation topic in health-care settings. Many medical schools have designed courses with the specific goal of cultivating empathy (Shapiro, Morrison, & Boker, 2004), and for good reason: Physician empathy has many benefits including both patient satisfaction with their care (Halpern, 2007) and physician job satisfaction (Gleichgerrcht & Decety, 2013). Low physician empathy, in contrast, predicts undertreatment of pain (Loewenstein, 2005). Despite its benefits, empathy can be costly and stressful. These costs apply generally—parental empathy is associated with stress-related inflammation (Manczak, DeLongis, & Chen, 2016)—and to health-care professionals specifically. Empathizing with life-and-death situations predicts fatigue and burnout (Gleichgerrcht & Decety, 2013), which is associated with worse physician–patient interactions (Dyrbye et al., 2010) and increased medical error (Shanafelt et al., 2012). Insofar as empathy impairs problem-solving (Haque & Waytz, 2012), inhibiting empathy may free up cognitive resources to treat patients more effectively (Decety, 2011). Thus, in many ways, the debate about physician empathy exemplifies how people must weigh empathy's costs and benefits and then choose whether to empathize (Zaki, 2014).

Cross-sectional work on self-reported empathy in physicians and nonphysicians has produced mixed results: Some studies suggest that physicians are higher in empathy (e.g., Handford, Lemon, Grimm, & Vollmer-Conna, 2013), others that physicians are *lower* (e.g., Decety, Yang, & Cheng, 2010), and still others that physicians and nonphysicians do not differ (e.g., Bellini & Shea, 2005). Moreover, longitudinal studies provide equivocal results for whether trait empathy decreases, increases, or remains stable over medical training (Neumann et al., 2011; Smith et al., 2017).

Moving beyond self-report measures, one neurophysiological study reported empathy differences between physicians and matched nonphysicians (Decety et al., 2010). Participants viewed images of hands being stuck with needles (normatively painful) or touched with Q-tips (normatively nonpainful). Neural activity previously associated with empathy was present among nonphysicians but not among physicians, suggesting that physicians may have blunted spontaneous empathy, which may help them avoid emotional exhaustion.

Our aim here was to complement this literature by investigating the underlying processes that shape physician empathy outcomes. This approach contrasts with previous approaches,

[1] Department of Psychology, The Pennsylvania State University, University Park, PA, USA
[2] Psychology Department, University of Kentucky, Lexington, KY, USA
[3] Psychology Department, University of California, Davis, CA, USA

Corresponding Author:
Victoria L. Spring, Department of Psychology, The Pennsylvania State University, 140 Moore Building, University Park, PA 16801, USA.
Email: vls23@psu.edu

which have typically focused on neural correlates of empathic processes and outcomes or self-reported empathic outcomes. Our focus on process affords a granular view of empathy as a construct and perhaps better prediction of empathic outcomes.

We used an implicit measure: the pain identification task (PIT; Cameron, Spring, & Todd, 2017). Participants view pairs of pain-relevant images (the same ones from Decety et al., 2010), and their focal task is to quickly judge whether the experiences depicted in the target images are painful or not, while avoiding influence of the experiences depicted in the preceding distracter images. Like other response-interference paradigms (e.g., Payne, 2001), participants display a robust interference effect, making more errors in pain judgments when distracter and target experiences are incongruent (painful/nonpainful and nonpainful/painful). Participants viewing a painful distracter and a nonpainful target, for example, might incorrectly judge the target experience as painful because they are still unintentionally empathizing with the affective content of the distracter experience.[1]

Rather than equating this behavioral effect with a single process, "empathy," we used multinomial modeling to dissociate latent processes underlying this effect. Multinomial modeling formally specifies the processes that interactively contribute to behavioral responses (Riefer & Batchelder, 1988). Observed behavior (i.e., accuracy) is used to estimate the probability each specified process is operating. Multinomial modeling has been applied to diverse topics including moral judgment (Cameron, Payne, Sinnott-Armstrong, Scheffer, & Inzlicht, 2017), prejudice (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005), visual perspective taking (Todd, Cameron, & Simpson, 2017), and, of most relevance here, empathy for pain (Cameron, Spring, & Todd, 2017).

The multinomial model for the PIT decomposes task performance into three processes: intentional empathy (IE: tendency to accurately judge targets as painful or nonpainful, consistent with participants' intention to accurately judge the target experience), unintentional empathy (UE: tendency to judge targets in a distracter-consistent manner, which occurs *despite* participants' intention to accurately judge the target experiences without being influenced by the distracter), and response bias (RB: a directional tendency to judge targets as painful or nonpainful). IE reflects empathy with target experiences: Participants following task instructions should focus on the (non)painfulness of target experiences and try to judge them accurately. If participants intentionally resonate with the target experiences, empathically sharing them, they should judge their painfulness more accurately. In contrast, participants engaging in UE should make a systematic pattern of errors in judging the target experiences because they are still resonating with the content of the distracter experiences. If participants unintentionally empathize with the distracter experiences, their judgments of the targets should be biased in the direction of the distracters. RB is a nonempathic process that may be implicated in accuracy rates as a tendency to judge targets as painful or nonpainful despite context.

These empathic processes, IE and UE, capture the experience-sharing facet of empathy. In contrast to mentalizing (taking another's perspective) or compassion (being motivated to reduce another's suffering), experience sharing reflects vicarious resonance with another's experiences (Decety & Cowell, 2014; Zaki & Ochsner, 2012). The PIT assesses both IE and UE, operationalizing them as follows: Whereas IE involves deliberately resonating with (and accurately detecting) the target experience, UE involves spontaneously resonating with the distracter experience, which biases target judgments in distracter-consistent way when IE fails. Furthermore, for UE, experience sharing with the distracter is counterintentional—it occurs despite participants' attempts to intentionally empathize only with the target experience. Our model assumes that UE drives responses when IE fails and that RB drives responses when the other two processes fail (Bishara & Payne, 2009).[2]

Cameron, Spring, & Todd (2017) used experimental manipulations to establish the construct validity of these process parameters. They found that imposing a fast response deadline reduced IE but not UE, suggesting that IE shares commonalities with resource-dependent controlled processes. UE was stronger when participants received self-focused (vs. other-focused) perspective-taking instructions, consistent with prior findings that such instructions increase spontaneous experience sharing (Jackson, Brunet, Meltzoff, & Decety, 2006). Because processes underpinning experience sharing of pain (e.g., negative affect and conceptual knowledge) also underpin nonempathic processes, Cameron, Spring, & Todd (2017) ruled out that empathy effects were reducible to affective or semantic priming. Reversing target pain outcomes—by stipulating that targets feel pain from Q-tips and no pain from needles—reduced IE and UE. Were effects driven by negative affect toward needles, outcome information should not have mattered. Additionally, on a control task wherein participants classified objects as needles or Q-tips rather than judging painfulness, this pain outcome manipulation had less influence. Were effects driven by accessible conceptual knowledge that needles are painful, this should not have happened. Thus, both forms of empathy captured by the PIT are likely not reducible to accuracy tracking.

We explored real-world implications of IE and UE, investigating them among practicing physicians and demographically matched nonphysicians. This is the first investigation to disentangle multiple forms of empathy within the same task to examine physician empathy. We attempted to address whether group differences in empathy are reducible to conceptual knowledge or negative affect by including control tasks assessing semantic priming (object identification task [OIT]) and affective priming (affective priming task [APT]). Because physicians have more experience observing pain, they might display differences in pain judgments simply because they are better at identifying pain-causing objects; with this greater experience, their basic affective experiences to depictions of pain may also differ. Overall, we expected that any differences between physicians and nonphysicians on the PIT would be reduced or eliminated in the control tasks.

Method

Participants

We recruited 40 practicing physicians and 40 nonphysicians matched on gender and education from the university community. Physicians held medical degrees; nonphysicians held comparable degrees in other fields. Sample size was determined before data collection based on available financial resources and time line to recruit participants before the end of July 2016. We excluded data from one physician who was an accuracy outlier (5% accuracy, suggesting response key confusion) and one control participant who did not complete all tasks, leaving a final sample of 39 physicians (19 men, 20 women; $M_{\text{age}} = 37.90$, $SD_{\text{age}} = 12.01$; 53.85% White, 23.08% Asian, 12.82% Other Ethnicity, 10.26% Latinx, and 2.56% Black, with some participants reporting more than one ethnicity) and 39 nonphysicians (19 men, 20 women; $M_{\text{age}} = 47.92$, $SD_{\text{age}} = 15.77$; 84.62% White, 12.82% Asian, 10.26% Latinx, 5.13% Black, and 2.56% Other Ethnicity, with some participants reporting more than one ethnicity). These exclusion rules, also established before data collection, are the same ones used by Cameron, Spring, & Todd (2017). A sensitivity analysis revealed that this sample size afforded 80% power to detect a medium-to-large effect ($\eta_p^2 = .10$). Physicians were younger than nonphysicians ($p = .002$); thus, we examined age effects on relevant outcome measures (see Tables S1–S3 in Supplemental Online Material [SOM]).

Procedure

Participants completed three sequential priming tasks in counterbalanced order: PIT, OIT, and APT. The latter two tasks aimed to distinguish empathy from semantic priming and nonempathic affective priming, respectively.

PIT. Participants viewed pairs of images in fast succession. They had to ignore the first image and quickly judge whether the second image depicted an experience that was painful or nonpainful. Distracter and target images depicted hands being stuck with needles or touched by Q-tips: experiences normatively judged as painful and nonpainful. There were two counterbalanced sets each of painful and nonpainful stimuli, with one set serving as distracters and the other as targets. The trial sequence was as follows: fixation cross (200 ms), distracter image (150 ms), blank screen (75 ms), and target image (until response). If participants did not respond within 400 ms, a warning appeared. Participants completed 15 trials of each distracter–target combination, resulting in 60 experimental trials (preceded by four practice trials).

OIT. The OIT was identical to the PIT, except participants judged whether the target image depicted a needle or a Q-tip. If the PIT assesses empathy, rather than pain assessment ability or semantic priming, we should expect reduced group differences in the OIT versus the PIT. Although PIT performance could capture some form of semantic knowledge, distinguishing this from other component processes of empathy may be difficult because empathy for pain requires some knowledge of the relevant outcomes. By including the OIT, we can test whether mere knowledge of painful outcomes explains empathy effects on the similarly structured PIT.

APT. The APT was identical to the PIT, except participants judged whether the target experience was “good” or “bad.” If the PIT assesses empathy rather than affective priming, we should expect reduced group differences in the APT versus the PIT. Negative affect and empathy for pain likely converge (Singer et al., 2004) because pain is typically experienced as bad. Nevertheless, by including the APT, we can test whether PIT performance is reducible to negative affect.

Exploratory measures. Participants completed several exploratory measures that are not discussed further. Details appear in SOM.

Results

Error Rates

We report results for each task separately. Table 1 displays error rates by condition.

PIT. A 2 (group) × 2 (distracter) × 2 (target) mixed analysis of variance (ANOVA) revealed a Distracter × Target interaction, $F(1, 76) = 26.82$, $p < .001$, $\eta_p^2 = .26$: Neutral targets were misidentified more often, and painful targets were misidentified less often, after painful (vs. nonpainful) distracters. The Group × Distracter × Target interaction was not significant ($F < 1$, $p = .536$[3]): Both physicians and nonphysicians exhibited the empathy interference effect. Given the age difference between groups, we tested whether age moderated these effects. There was a significant Age × Distracter × Target interaction, $F(1, 76) = 16.31$, $p < .001$, $\eta_p^2 = .18$: The empathy interference effect was stronger for younger participants. Table S1 in SOM reports descriptive statistics at younger (≤37 years) and older (>37 years) age strata to complement the stratification analyses reported below.

OIT. An identical ANOVA on the OIT also revealed a Distracter × Target interaction, $F(1, 76) = 26.27$, $p < .001$, $\eta_p^2 = .26$: Q-tip targets were misidentified more often, and needle targets were misidentified less often, after needle (vs. Q-tip) distracters. The Group × Distracter × Target interaction was not significant ($F < 1$, $p = .627$). As with the PIT, there was an Age × Distracter × Target interaction, $F(1, 76) = 6.16$, $p = .015$, $\eta_p^2 = .08$: The interference effect was stronger for younger participants (Table S1 in SOM displays OIT results for each age strata).

Table 1. Mean Proportion of Errors by Participant Group, Task Type, Distracter Type, and Target Type.

	Participant Group							
	Physicians				Nonphysicians			
Task/Target	Painful/Needle Distracter	Neutral/Q-Tip Distracter	95% CI	Hedges' g_{av}	Painful/Needle Distracter	Neutral/Q-Tip Distracter	95% CI	Hedges' g_{av}
PIT								
Painful	.06 (.10)	.12 (.16)	[.03, .10]	.47	.09 (.14)	.17 (.22)	[.03, .13]	.40
Neutral	.13 (.19)	.05 (.07)	[.03, .13]	.54	.13 (.19)	.09 (.15)	[−.00, .08]	.22
OIT								
Needle	.05 (.07)	.13 (.15)	[.03, .12]	.62	.04 (.07)	.13 (.15)	[.05, .13]	.72
Q-tip	.12 (.20)	.04 (.07)	[.02, .14]	.54	.10 (0.13)	.05 (.09)	[.01, .08]	.37
APT								
Needle	.05 (.06)	.14 (.15)	[.04, .13]	.74	.09 (0.16)	.14 (.19)	[.01, .09]	.27
Q-tip	.16 (.21)	.06 (.08)	[.04, .16]	.60	.12 (0.17)	.11 (.16)	[−.02, .04]	.08

Note. Standard deviations are in parentheses. Ninety-five percent CIs are for incongruent–congruent mean difference. For the PIT, all simple effects, $p < .005$, except for neutral targets among nonphysicians ($p = .058$). For the OIT, all simple effects, $p \leq .020$. For the APT, all simple effects, $p < .012$, except for neutral targets among nonphysicians ($p = .372$). PIT = pain identification task; OIT = object identification task; APT = affective priming task; CI = confidence interval.

Table 2. Parameter Estimates by Task and Participant Group.

	Physicians		Nonphysicians		Across Groups	
Task/Parameter	Estimate [95% CI]	$\Delta G^2(1)$	Estimate [95% CI]	$\Delta G^2(1)$	$\Delta G^2(1)$	*w*
PIT						
IE	.82 [.79, .84]	1,837.64**	.76 [.74, .79]	1,557.19**	8.92**	.04
UE	.39 [.27, .51]	37.32**	.23 [.13, .34]	17.32**	3.73†	.03
RB	.48 [.38, .58]	.14	.45 [.38, .52]	1.64	.20	.01
OIT						
D	.83 [.81, .85]	1,911.66**	.84 [.82, .86]	1,959.64**	.41	.01
AB	.46 [.34, .58]	48.20**	.40 [.28, .53]	33.73**	.41	.01
RB	.45 [.34, .57]	.64	.50 [.39, .61]	.00	.31	.01
APT						
IV	.79 [.77, .82]	1,723.54**	.77 [.74, .80]	1,577.75**	1.62	.02
UV	.44 [.34, .55]	55.25**	.14 [.03, .25]	5.77*	14.61**	.06
RB	.54 [.44, .64]	.65	.50 [.44, .57]	.01	.38	.01

Note. The neutral reference point for IE/D/IV and UE/AB/UV is 0, and the neutral reference point for RB is .50. PIT = pain identification task; OIT = object identification task; APT = affective priming task; CI = confidence interval. In the PIT, IE = intentional empathy, UE = unintentional empathy, and RB = response bias. In the OIT, D = discriminability, AB = accessibility bias, and RB = response bias. In the APT, IV = intentional valence, UV = unintentional valence, and RB = response bias.
†$p < .100$. *$p < .050$. **$p < .010$.

APT. An identical ANOVA on the APT also yielded a Distracter × Target interaction, $F(1, 76) = 18.88$, $p < .001$, $\eta_p^2 = .20$: The valence of Q-tip targets was misjudged more often, and the valence of needle targets was misjudged less often, after needle (vs. Q-tip) distracters. Unlike the other tasks, however, there was a significant Group × Distracter × Target interaction, $F(1, 76) = 4.53$, $p = .037$, $\eta_p^2 = .06$: The Distracter × Target interaction was stronger among physicians, $F(1, 38) = 14.12$, $p = .001$, $\eta_p^2 = .27$, than nonphysicians, $F(1, 38) = 4.77$, $p = .035$, $\eta_p^2 = .11$. Physicians misjudged the valence of both needle and Q-tip targets more often after incongruent distracters. As before, there was an Age × Distracter × Target interaction, $F(1, 76) = 7.35$, $p = .008$, $\eta_p^2 = .09$: The interference effect was stronger for younger participants (Table S1 in SOM presents APT results for each age strata).

Multinomial Models

Next, we conducted multinomial modeling analyses for all three tasks using MultiTree (Moshagen, 2010). Although the models were similar across tasks, the conceptual meaning of the parameters differs by task (see below). We estimated parameters for each participant group and then tested whether constraining each parameter to be equivalent across groups significantly reduced model fit (ΔG^2). Table 2 displays parameter estimates for each task by group. Because physicians were younger than nonphysicians, for each model, we also conducted stratified analyses to test whether group differences in the parameters differed by age (Tables S2 and S3 in SOM). The process model and equations for all three tasks appear in SOM.

PIT. The PIT model estimated three underlying processes: IE (ability to accurately judge the painfulness of target experiences), UE (tendency to judge the painfulness of target experiences in a distracter-consistent manner, capturing spontaneous experience sharing with distracter experiences), and RB (directional tendency to always judge target experiences as painful). UE drives responses when IE fails; RB drives responses when both IE and UE fail.

Initial analysis. We estimated IE, UE (constrained to be equal across pain-distracter and neutral-distracter conditions), and RB parameters for each participant group and compared these estimates against zero (or chance [.50], for RB) and across groups. We examined change in model fit (ΔG^2) when imposing such constraints and report effect size w, with $w \leq .05$ indicating adequate model fit (Clerkin, Fisher, Sherman, & Teachman, 2014). The model fit the data, $G^2(2) = 2.34$, $p = .310$, $w = .02$. Physicians were higher in IE ($p = .003$) and marginally higher ($p = .053$) in UE than nonphysicians, whereas RB did not significantly differ across groups.

Stratified analysis. Because age differed across groups and moderated PIT effects, nonphysicians might have had lower IE because they were older. Prior work has found a negative association between age and the control parameter in process dissociation (analogous to the IE parameter here and representing the ability to accurately judge target content), possibly reflecting age-related declines in inhibitory ability (Stewart, von Hippel, & Radvansky, 2009). Because MultiTree cannot conduct moderation analyses to identify covariates, we used stratification analysis to examine potential age effects. Stratification analysis tests the relationship between variables while holding other variables constant, splitting the age variable into multiple groups (Greenland & Rothman, 1998). This technique, used in public health (e.g., Backhans & Hemmingsson, 2012; Rosen et al., 2004) and psychological research (e.g., Fergusson & Horwood, 1995) to account for demographic moderators, entails finding cut points on a moderating covariate (e.g., age) to create strata and examining relationships within each stratum.

This procedure produced a “younger” stratum ($\leq$37 years; 27 physicians, 13 nonphysicians) and an “older” stratum (>37 years; 12 physicians, 26 nonphysicians). The groups did not significantly differ in age in either age stratum (ps $\geq$ .45). We estimated an identical model in each stratum. The model fit the data in both—younger: $G^2(2) = 3.82$, $p = .148$, $w = .04$; older: $G^2(2) = 0.74$, $p = .691$, $w = .02$. In both age strata, IE was higher for physicians, whereas UE did not significantly differ between groups.

These results suggest the initial effect, whereby physicians exhibit higher IE than nonphysicians, is not reducible to age-related differences in controlled processing. Physicians displayed higher IE regardless of age strata, but the group difference was *stronger* in the older group ($w = .11$) versus the younger group ($w = .05$), perhaps reflecting motivation to intentionally empathize with patients' pain or physicians' expertise in pain identification. The stratified analyses also suggest the initial marginal group effect on UE may have been moderated by age differences between groups. The lack of a significant group effect on UE contrasts with prior findings that physicians exhibit less spontaneous empathy than do nonphysicians (Decety et al., 2010).

OIT. The OIT model estimated discriminability (D: ability to accurately categorize target objects as needles or Q-tips), accessibility bias (AB: distracter-consistent activation of knowledge about needles and Q-tips elicited by distracter stimuli), and RB (directional tendency to classify target objects as needles). The conceptual meaning of the parameters differs, but the models for the OIT and PIT are structurally identical.

Initial analysis. We estimated D, AB (constrained to be equal across needle-distracter and Q-tip-distracter conditions), and RB parameters for each participant group. The model fit the data, $G^2(2) = 4.07$, $p = .131$, $w = .03$. The groups did not differ on any parameters. If IE were reducible to semantic priming or expertise at classification, we should see a similar pattern of results for D by group, with D (like IE) being higher among physicians. Instead, the groups displayed comparable ability to discriminate between objects, which suggests that IE is not reducible to accessible semantic knowledge.

Stratified analysis. As with the PIT, we estimated an identical model in each age stratum. In the younger stratum, the model fit the data, $G^2(2) = 2.12$, $p = .347$, $w = .03$. Neither D nor AB significantly differed between groups. In the older stratum, the model fit the data, $G^2(2) = 2.86$, $p = .240$, $w = .04$, and D was stronger among physicians. Because the group effect on D only emerged in the older stratum, we refrain from further interpretation.

APT. The APT model estimated intentional valence (IV: ability to accurately judge the affective valence of target experiences as good or bad), unintentional valence (UV: tendency to judge the valence of target experiences in a distracter-consistent manner), and RB (directional tendency to judge target experiences as negative). The model is structurally identical to those for the PIT and OIT, but the conceptual meaning of the parameters differs.

Initial analysis. We estimated IV, UV (constrained to be equal across needle-distracter and Q-tip-distracter conditions), and RB parameters for each participant group. The model fit the data, $G^2(2) = 2.16$, $p = .339$, $w = .02$. Neither IV nor RB significantly differed between groups, but UV was stronger among physicians. Therefore, it seems that group differences in UV underlie the Group $\times$ Distracter $\times$ Target interaction on the APT. That no group difference in UE emerged on the PIT for both age strata suggests the PIT and APT tap different processes. Unlike in the PIT model, wherein IE was stronger for physicians, IV in the APT did not significantly differ between groups. If both IE and UE were reducible to affective priming, we would expect a similar pattern of results for IV and UV by group. Instead, the opposite pattern of results emerged for physicians versus nonphysicians, which suggests the IE parameter may reflect a process that is not reducible to affective priming.

Stratified analysis. As before, we estimated an identical model in each age stratum. In the younger stratum, the model fit the data, $G^2(2) = 1.35, p = .510, w = .02$. IV did not significantly differ between groups, but UV was stronger among physicians. In the older stratum, the model fit the data, $G^2(2) = 1.01, p = .603, w = .02$. IV was stronger among physicians, but UV did not significantly differ between groups. Because the group effects on IV and UV did not emerge in both age strata, we do not interpret them further. We note, however, that the group effect on the valence parameters was less stable as a function of age, perhaps reflecting a role of experience in how bad one feels when viewing painful experiences. This finding contrasts with the PIT, for which group differences in IE but not UE emerged in both age strata, providing further evidence that processes underlying PIT performance may not be reducible to negative affect.

Discussion

The paradox of physician empathy has motivated much study across psychology and medicine: Physicians aim to understand and care for their patients, yet they also must maintain emotional objectivity. Some prior work suggests that physicians display reduced empathy on self-report and neurophysiological measures, whereas other work suggests negligible empathy differences. We used implicit measurement and multinomial modeling to understand physicians' intentional and unintentional empathy for pain. Relative to matched nonphysicians, physicians were more likely to intentionally empathize with others' pain but, contrary to some prior work, were no less likely to experience spontaneous empathy.

We replicated prior findings that people mistakenly judge the painfulness of target experiences after distracter experiences depicting incongruent pain content (Cameron, Spring, & Todd, 2017). Although this behavioral effect was not moderated by participant group, multinomial modeling analyses unmasked a group difference in the ability to accurately judge the painfulness of target experiences. This result reveals the utility of a formal modeling approach: By mathematically disentangling, and specifying a priori, the processes underpinning task performance, multinomial modeling identifies relationships that might otherwise go undetected if focusing only on behavioral outcomes (Payne, 2008).

Importantly, the difference in IE between physicians and nonphysicians was not reducible to age differences between groups. Because nonphysicians were older than physicians, they may have displayed lower IE due to age-related declines in cognitive control. However, stratification analyses revealed that once accounting for group differences in age, the group difference in IE became even stronger among older participants. One possible explanation is that with age, physicians gain expertise in accurately diagnosing others' pain. This result aligns with recent findings that certain types of empathy (i.e., trait perspective taking and trait empathic concern) may increase during medical school (Smith et al., 2017).

Additionally, the group difference in IE was not reducible to domain general semantic or affective priming. Although knowledge about painful outcomes is likely implicated in empathy for pain, physicians did not display greater D than nonphysicians on the OIT. Similarly, on the APT, physicians and nonphysicians did not differ in the ability to accurately judge target experiences as good or bad, suggesting the group effect on IE in the PIT does not merely reflect differences in negative affective responses to harmful stimuli. Instead, intentionally judging the painfulness of target experiences seemed to involve the ability to understand through experience sharing which types of experiences are painful for others. Because IE reflects empathy with the target experience, insofar as participants experience share with the target, they should be more accurate in judging the target experience's painfulness. Thus, intentionally resonating with the experiential content of the target should improve accuracy when judging that target.

Group differences emerged for UV (i.e., negative affective responses to distracters) but only marginally for UE. UE is the parameter most comparable to empathy as measured in past neurophysiological studies of empathy for pain among physicians (Decety et al., 2010), so this null effect is noteworthy. In summary, physicians in our study outperformed nonphysicians in intentionally empathizing with others' pain, independent of their ability to categorize objects or make valence judgments.

Our IE findings complement past work suggesting physicians experience empathy differently, and possibly more strongly (e.g., Handford et al., 2013), than nonphysicians. However, our UE results are inconsistent with prior work suggesting physicians show reduced spontaneous empathy on neurophysiological measures in response to the same stimuli (Decety et al., 2010). These distinct effects reveal the utility of using modeling to disentangle and quantify IE and UE for pain.

What separates physicians from nonphysicians may be the former group's greater likelihood of intentionally empathizing when it is helpful to do so. Although our task cannot address questions about the normative benefits of empathy in medical practice, it provides a formalized description of how physicians experience empathy for pain. It is also an open question whether group differences in IE are attributable to career selection (people higher in IE choosing to become physicians) or to experience and exposure (expertise developed as a practicing physician leading to increased IE). The age effects—higher IE among older versus younger physicians—afford some speculation. Physicians in the younger age stratum exhibited higher IE than nonphysicians, which may suggest the existence of a selection effect. People with higher ability to intentionally empathize may be more likely to choose to become physicians. Physicians in the older age stratum, however, displayed an even greater increase in IE compared with nonphysicians. It is possible that greater experience and expertise among physicians amplifies their tendency to engage in IE relative to nonphysicians. Future work should examine this possibility directly using a longitudinal design.

These results also speak theoretically to the nature of empathy. Our findings are broadly consistent with motivated accounts of empathy (Keysers & Gazzola, 2014; Zaki, 2014). Such accounts posit that people actively choose to empathize or not depending on whether empathy aligns with their current goals, which may include the goal of minimizing anticipated emotional exhaustion (Cameron, Harris, & Payne, 2016). It is possible that experience with painful situations requires more emotional regulation to cope with the exhaustion from empathizing in such situations (Cheng et al., 2007), which may explain why we observed an increased tendency among people with expertise in treating pain to engage in intentional (but not unintentional) empathy for pain.

Understanding how empathy operates among physicians is relevant not only to the study of clinical empathy but also to the study of empathy in general. We explored the relationship between relatively automatic (unintentional) and relatively controlled (intentional) forms of empathy for pain. The automatic-controlled distinction is a core concern in discussions about empathy (de Vignemont & Singer, 2006; Hodges & Wegner, 1997); our data speak directly to this issue. We can also extrapolate from these results to speculate about the role of empathy regulation and empathy fatigue more broadly. For example, our findings suggest that intentional types of empathy may be more easily regulated than spontaneous forms of empathy. An informative next step could be to manipulate task instructions to investigate whether people intentionally shift empathy to track with changing instructions. It may also be that regulating intentional empathy improves resistance to burnout and fatigue and that expertise with painful situations plays a role in one's ability to regulate intentional (but not unintentional) empathy for pain. Our findings lay the groundwork for these future directions.

Our findings also highlight the importance of attending to idiographic factors such as expertise and experience when measuring empathy in normal populations. By examining how IE and UE vary across populations with different levels of empathy and pain-relevant expertise, we can begin to understand idiographic variability in these distinct types of empathy.

This work has several limitations. Due to the difficulty in recruiting such participant populations, sample size was low and participants were not matched on age. Thus, we were underpowered to detect smaller differences in UE between groups. Low sample size can increase the likelihood of both false-positive and false-negative results (Button et al., 2013). Although we maintain that our stratification analyses accounted for the moderating effect of age, future work examining empathy among physicians and nonphysicians should match participants on age during recruitment. Future research might also investigate IE among physicians using longitudinal designs to determine whether the age differences observed here reflect cohort effects or developmental changes across one's career. Another avenue for future work is to examine differences in IE and UE across different medical specialties (e.g., surgery and psychiatry) or professions (e.g., physicians and nurses) and to establish convergent validity by examining correlations between the empathic processes measured here and other assessments of empathy. One limitation of such an approach, however, is that past work has often found low correlations between self-reported trait empathy and empathy as measured by implicit methods and neurophysiology (Cheng et al., 2007; Lamm, Meltzoff, & Decety, 2010).

Overall, these findings suggest physicians are likely to empathize with others' pain when it aligns with their overt goals but may be no more (or less) likely than nonphysicians to spontaneously empathize with pain. By using implicit measurement and formal modeling, we evade some of the limitations of self-report and provide an effective and efficient way of dissociating individual differences in empathy among physicians.

Acknowledgments

We thank Astrid Montuclard and Vincent Peiffer for research assistance and Jatin Vaidya for providing laboratory testing resources.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was facilitated by an NSF Graduate Research Fellowship (awarded to V.L.S.), a University of Iowa Graduate and Professional Student Government Research Grant (awarded to V.L.S.), and NSF Grants BCS-1660707 (awarded to C.D.C.) and BCS-1523731 (awarded to A.R.T.).

Supplemental Material

The supplemental material is available in the online version of the article.

Notes

1. We focus on a type of empathy that does not purely reflect pain assessment but rather one that also involves vicariously sharing the affective experience of pain (Singer et al., 2004; Zaki, Wager, Singer, Keysers, & Gazzola, 2016). We are agnostic about whether participants are experiencing physical pain when resonating with the images.
2. We used an intentional empathy (IE)–dominating model to be consistent with previous work using this task (Cameron, Spring, & Todd, 2017) and similar sequential-priming tasks (Bishara & Payne, 2009), assuming that counterintentional empathy would only occur when intentional empathy failed. This assumption aligns with our conceptual definitions of intentional and unintentional empathy.
3. For all nonsignificant effects, η_p^2 s < .01, unless otherwise noted.

References

Backhans, M. C., & Hemmingsson, T. (2012). Unemployment and mental health—Who is (not) affected? *European Journal of Public Health*, *22*, 429–433.

Bellini, L. M., & Shea, J. A. (2005). Mood change and empathy decline persist during three years of internal medicine training. *Academic Medicine*, *80*, 164–167.

Bishara, A. J., & Payne, B. K. (2009). Multinomial process tree models of control and automaticity in weapon misidentification. *Journal of Experimental Social Psychology*, *45*, 524–534.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews: Neuroscience*, *14*, 365–376.

Cameron, C. D., Harris, L. T., & Payne, B. K. (2016). The emotional cost of humanity: Anticipated exhaustion motivates dehumanization of stigmatized targets. *Social Psychological and Personality Science*, *7*, 105–112.

Cameron, C. D., Payne, B. K., Sinnott-Armstrong, W., Scheffer, J. A., & Inzlicht, M. (2017). Implicit moral evaluations: A multinomial modeling approach. *Cognition*, *158*, 224–241.

Cameron, C. D., Spring, V. L., & Todd, A. R. (2017). The empathy impulse: A multinomial model of intentional and unintentional empathy for pain. *Emotion*, *17*, 396–411.

Cheng, Y., Lin, C. P., Liu, H. L., Hsu, Y. Y., Lim, K. E., Hung, D., & Decety, J. (2007). Expertise modulates the perception of pain in others. *Current Biology*, *17*, 1708–1713.

Clerkin, E. M., Fisher, C. R., Sherman, J. W., & Teachman, B. A. (2014). Applying the quadruple process model to evaluate change in implicit attitudinal responses during therapy for panic disorder. *Behaviour Research and Therapy*, *52*, 17–25.

onrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The quad-model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487.

Decety, J. (2011). The neuroevolution of empathy. *Annals of the New York Academy of Sciences*, *1231*, 35–45.

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science*, *9*, 525–537.

Decety, J., Yang, C. Y., & Cheng, Y. (2010). Physicians down-regulate their pain empathy response: An event-related brain potential study. *NeuroImage*, *50*, 1676–1682.

de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, *10*, 435–441.

Dyrbye, L. N., Thomas, M. R., Power, D. V., Durning, S., Moutier, C., Massie, F. S., Jr., . . . Shanafelt, T. D. (2010). Burnout and serious thoughts of dropping out of medical school: A multi-institutional study. *Academic Medicine*, *85*, 94–102.

Fergusson, D. M., & Horwood, L. J. (1995). Early disruptive behavior, IQ, and later school achievement and delinquent behavior. *Journal of Abnormal Child Psychology*, *23*, 183–199.

Gleichgerrcht, E., & Decety, J. (2013). Empathy in clinical practice: How individual dispositions, gender, and experience moderate empathic concern, burnout, and emotional distress in physicians. *PLoS One*, *8*, e61526.

Greenland, S., & Rothman, K. J. (1998). Introduction to stratified analysis. In K. Rothman, S. Greenland, & T. L. Lash (Eds.), *Modern epidemiology* (Vol. 3, pp. 248–304). Philadelphia, PA: Lippincott, Williams, & Wilkins.

Halpern, J. (2007). Empathy and patient-physician conflicts. *Journal of General Internal Medicine*, *22*, 696–700.

Handford, C., Lemon, J., Grimm, M. C., & Vollmer-Conna, U. (2013). Empathy as a function of clinical exposure—Reading emotion in the eyes. *PLoS One*, *8*, e65159.

Haque, O. S., & Waytz, A. (2012). Dehumanization in medicine: Causes, solutions, and functions. *Perspectives on Psychological Science*, *7*, 176–186.

Hodges, S., & Wegner, D. (1997). Automatic and controlled empathy. In W. J. Ickes (Ed.), *Empathic accuracy* (pp. 311–339). New York, NY: Guilford Press.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia*, *44*, 752–761.

Keysers, C., & Gazzola, V. (2014). Dissociating the ability and propensity for empathy. *Trends in Cognitive Sciences*, *18*, 163–166.

Lamm, C., Meltzoff, A. N., & Decety, J. (2010). How do we empathize with someone who is not like us? A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *22*, 362–376.

Loewenstein, G. (2005). Hot-cold empathy gaps and medical decision making. *Health Psychology*, *24*, S49–S56.

Manczak, E. M., DeLongis, A., & Chen, E. (2016). Does empathy have a cost? Diverging psychological and physiological effects within families. *Health Psychology*, *35*, 211–214.

Moshagen, M. (2010). Multitree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42–54.

Neumann, M., Edelhäuser, F., Tauschel, D., Fischer, M. R., Wirtz, M., Woopen, C., . . . Scheffer, C. (2011). Empathy decline and its reasons: A systematic review of studies with medical students and residents. *Academic Medicine*, *86*, 996–1009.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–192.

Payne, B. K. (2008). What mistakes disclose: A process dissociation approach to automatic and controlled processes in social psychology. *Social and Personality Psychology Compass*, *2*, 1073–1092.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318–339.

Rosen, C. L., Storfer-Isser, A., Taylor, H. G., Kirchner, H. L., Emancipator, J. L., & Redline, S. (2004). Increased behavioral morbidity in school-aged children with sleep-disordered breathing. *Pediatrics*, *114*, 1640–1648.

Shanafelt, T. D., Boone, S., Tan, L., Dyrbye, L. N., Sotile, W., Satele, D., . . . Oreskovich, M. R. (2012). Burnout and satisfaction with work-life balance among US physicians relative to the general US population. *Archives of Internal Medicine*, *172*, 1377–1385.

Shapiro, J., Morrison, E., & Boker, J. (2004). Teaching empathy to first year medical students: Evaluation of an elective literature and medicine course. *Education for Health*, *17*, 73–84.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*, 1157–1162.

Smith, K. E., Norman, G. J., & Decety, J. (2017). The complexity of empathy during medical school training: evidence for positive changes. *Medical Education*, *51*, 1146–1159.

Stewart, B. D., von Hippel, W., & Radvansky, G. A. (2009). Age, race, and implicit prejudice using process dissociation to separate the underlying components. *Psychological Science*, *20*, 164–168.

Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin*, *140*, 1608–1647.

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: Progress, pitfalls and promise. *Nature Neuroscience*, *15*, 675–680.

Zaki, J., Wager, T. D., Singer, T., Keysers, C., & Gazzola, V. (2016). The anatomy of suffering: Understanding the relationship between nociceptive and empathic pain. *Trends in Cognitive Sciences*, *20*, 249–259.

Author Biographies

Victoria L. Spring is a PhD student in the Psychology Department at the Pennsylvania State University. Her research focuses on empathy, moral emotions, and moral judgment.

C. Daryl Cameron is an assistant professor of psychology and research associate in the Rock Ethics Institute at Penn State University. His research focuses on motivational and implicit processes that shape empathy and moral decision-making.

Stephanie McKee is a PhD student in the Psychology Department at the University of Kentucky.

Andrew R. Todd is an assistant professor of psychology at the University of California, Davis. His research focuses on empathy, perspective taking, and mental state reasoning.

Handling Editor: Joseph Cesario