Appendix A: Table 1 from Quigley & Barrett, 2014 (Reprinted with Permission)

Table 1
Meta-analytic/qualitative review summaries for specificity outcomes.

	Anger	Fear	Happiness	Sadness	Disgust
Cacioppo et al. (2000)	>↑ in HR rel. to disgust and happy; <↑ in HR rel. to fear	>↑ in HR rel. to anger, disgust, happy and sad	<↑ in HR rel. to anger	>↑ in HR rel. to disgust	<↑ in HR rel. to anger, fear and sad; HR not diff. from control
	>↑ in DBP rel. to fear, happy, and sad	<↑ in DBP rel. to anger	<↑ in DBP rel. to anger and sad	<↑ in DBP rel. to anger; >↑ in DBP rel. to happy	
	>NNSCRs rel. to fear	>NNSCRs rel. to sad		<NNSCRs rel. to fear	
		<↑SCL rel. to sad	<↑SCL rel. to disgust	>↑SCL rel. to fear	>↑SCL rel. to happy
	>↑ TPR and finger PVA and <↑ in CO and SV rel. to fear	<↑ TPR and finger PVA and >↑ in CO and SV rel. to anger			
	>↑ in face temp rel. to fear	<↑ in face temp rel. to anger			
Stemmler (2004)[a]	>↑ in DBP rel. to fear	<↑ in DBP rel. to anger			
	>↑ in TPR rel. to fear	<↑ in TPR rel. to anger			
	<↑ in CO rel. to fear	>↑ in CO rel. to anger			
	>↑ in facial temp rel. to fear	<↑ in facial temp rel. to anger			
	<↑ in resp. rate rel. to fear[b]	>↑ in resp. rate rel. to anger[b]			
Kreibig (2010)[c]	↑ in HR	↓ in HR (IT)	↑ in HR	↓ in HR (NC)	↓ in HR (MT)
		↑ in HR		↓ in HR (NC)	↓ in HR (MT)
				↑ in HR (CR)	↓ in HR (MT)
	↓ in HRV	↓ in HRV	↓ in HRV	↓ in HRV (NC)	↑ in HRV (CT)
	↑ NNSCR	↑ NNSCR	↑ NNSCR	↓NNSCR(NC)	↑ NNSCR (CT and MT)
	↑ in SCL	↓ in SCL (IT)		↓ in SCL (NC)	↑ in SCL (CT and MT)
		↑ in SCL		↓ in SCL (NC)	
	↑ in TPR	↓ in TPR		↑ in TPR (ANT)	↑ in TPR (CT)
	↓ in FPTT	↓ in FPTT	↑ in FPTT		
		↓ in EPTT	↑ in EPTT		
	↓ in face temp	↓ in face temp	↑ in face temp	↓ in face temp (CR and NC)	
		↓ in Expir. Time	↓ in Expir. Time		↑ in Expir. Time (CT)
				↑ in Tidal Volume (NC)	↓ in Tidal Volume (CT)

Notes. Variables are usually reported as change from a neutral control or from baseline. >↑ = greater increase; <↑ = smaller increase; >#= greater number; <#= fewer. We report here on the most commonly reported ANS measures for clarity.

[a] Stemmler (2004) compared only anger and fear; and for this meta-analysis we report only those effect sizes in which the weighted mean effect size was significant.

[b] In Stemmler, 2004, respiration rate was the only outcome measure for which the heterogeneity in effect sizes across studies was NOT significant.

[c] The Kreibig review did not compare specific emotions to one another, but rather proposed a modal ANS response pattern for each emotion. Thus, we can distinguish here only emotions for which the physiological responses across two emotions had a modally different direction of change (see Kreibig's Table 2); we included here only cases where the modal response direction was observed in a minimum of 3 studies. For heart rate (HR), multiple rows are shown to distinguish responses that differentiated emotion subforms (per Kreibig), and comparisons are made within a row. Emotion subforms are indicated by the codes in parentheses as follows: for fear, (IT) = imminent threat vs. typical fear (no abbreviation); for sadness, (CR) = crying or (NC) = non-crying or (ANT) = anticipatory; for disgust, = (CT) = contamination or (MT) = mutilation.

Physiological abbreviations used in the table: CO = cardiac output; DBP = diastolic blood pressure; EPTT = ear pulse transit time; Expir. Time = respiratory expiration time; facial temp = facial temperature; finger PVA = finger pulse volume amplitude; FPTT = finger pulse transit time; HR = heart rate; HRV = heart rate (or heart period) variability; NNSCR = number of non-specific skin conductance responses; SCL = skin conductance level; SV = stroke volume; TPR = total peripheral resistance.

Note. This table is reprinted with permission from Quigley & Barrett (2014). Compare the interpretation of the Stemmler (2004) meta-analysis in relation to a meta-analysis by Cacioppo and colleagues a few years earlier (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000). Cacioppo et al. (2000, 22 studies) observed

greater increases in diastolic blood pressure (DBP) and heart rate (HR) in studies of anger relative to studies of fear. Cacioppo et al. (2000) were careful to note the low consistency in the effect sizes across studies, and inferred that this variability made it difficult to interpret the pattern of findings as evidence of anger and fear fingerprints, ultimately concluding that there was relatively little evidence of emotion specificity across the 22 studies they examined (Cacioppo et al., 2000). Stemmler (2004) assessed consistency and specificity in the autonomic patterns of 15 studies that examined anger and fear categories specifically (8 of which also appeared in the (Cacioppo et al., 2000 meta-analysis). Stemmler (2004) replicated the Cacioppo et al. (2000) DBP finding, but failed to replicate the HR finding. Stemmler (2004) also identified mean differences in respiration rate (RR), total peripheral resistance (TPR), cardiac output (CO) and facial temperature when comparing anger and fear studies, but this was in the face of substantial variation (i.e., little evidence of consistency) for all measures except RR. However, Stemmler (2004) concluded that he had found evidence for autonomic fingerprints for anger and fear, while also indicating that there were notable effects of context. In fact, if we consider what was taken as evidence of specificity (mean differences across emotion categories) in the context of the absence of evidence for consistency (substantial variation across studies within one emotion category), and inconsistent findings across meta-analyses, these meta-analyses constitute weak support for the emotion fingerprints hypothesis (assuming methodological explanations are ruled out) and is the kind of evidence that would provide support for the emotion populations hypothesis.

Appendix B: Operationalization of Potential Moderators

Moderator	Value	Coding description and criteria
Sample Moderators		
Sample Type	1 = College Sample 2 = Community Sample	Categorical variable representing whether the sample consisted of university students or members of the community.
Gender	1 = All women 2 = Mixed sample 3 = All men	Categorical variable representing the gender breakdown of the experimental sample.
Study Moderators		
Induction Type	1 = Imagery & Recall 2 = Movies 3 = Images & Faces 4 = Social 5 = Sounds & Music 6 = Threat	Categorical variable representing the type of procedure used to induce emotion. (See, Supplemental Table 1 for a detailed description of each induction type)
Evaluation	0 = No evaluation 1 = Evaluation	Categorical variable representing whether participants were being evaluated as part of the experimental task. In order for a study to be marked as containing evaluation, participants needed to be aware that they were being evaluated (e.g. participants completing a mental math task were given feedback when they made a mistake).
Video Recorded	0 = No video recorded 1 = Video recorded	Categorical variable representing whether participants were being video recorded during the experimental task. In order for a study to be marked as being video recorded, participants needed to be aware that they were being recorded (e.g. participants practiced a speech in front of a video camera or experimenters prompted participants during a recall task from another room)
Presence of Others	0 = Alone 1 = Others present	Categorical variable representing whether participants were alone during the experimental task (e.g. watching a movie in an experimental testing room) or in the presence of others (e.g. engaging in a debate with a confederate)
Emotion Words	0 = No emotion words 1 = Emotion words	Categorical variable representing the presence or absence of emotion words during the experiment, either as part of the manipulation (e.g. participants were prompted to

		describe a time when they felt angry or fearful) or if emotion words were interspersed throughout the task (e.g. participants rated how much they were feeling certain emotions in between blocks of stimuli). Studies *were not* designated as containing emotion words if emotion words were included at the end of the study (e.g. a post-experiment manipulation check).
Quality Moderators		
Baseline Duration	1 = Less than a minute 2 = 1-5 minutes 3 = 5-10 minutes 4 = More than 10 minutes	Ordinal variable representing the length of the baseline period during physiological recording. Baseline duration was coded as the amount of time of that the initial baseline, prior to the emotion induction task(s).
No. Emotions Induced	1 = Single emotion 2 = Two emotions 3 = Three emotions 4 = Four emotions 5 = Five emotions 6 = Greater than five	Ordinal variable representing the number of unique emotions induced in the same participant during the experiment (e.g. If participants watched a video to induce anger, then a watched video to induce fear, this study was coded as 2). Neutral emotion inductions were included in this variable.
Induction Duration	1 = Less than a minute 2 = 1-5 minutes 3 = 5-10 minutes 4 = More than 10 minutes	Ordinal variable representing the length of each emotion induction. Induction duration was coded as the amount of time that a single emotion was induced. For example, if disgust was induced using a block of 20 images shown for 6-seconds, it was coded as a 2 because the length of the disgust induction was 120 seconds (2 minutes). We coded it a 2 even if there were multiple disgust blocks in the experiment (e.g. five disgust blocks were interspersed with five neutral blocks) because a single emotion induction was 120 seconds long.
Manipulation Check	0 = No manipulation check 1 = Manipulation check	Categorical variable representing whether (or not) the study included a check that the manipulation successfully induced an instance of the target emotion. Manipulation check was only coded as present if manipulation check was statistically significant (i.e. emotion was successfully induced).
Rigor of Induction	1 = Low 2 = Moderate 3 = High	Categorical variable representing the rigor of the induction procedure. This was an "inferential" variable measuring the coder's sense of the methodological rigor of the induction procedure. Coder instructions were as follows: "*Coder, this is an "inferential" variable. Many of the components that objectively make up "rigor" are recorded elsewhere. Please rate your "sense" of the methodological rigor of the*

		induction procedure. Some possible aspects to consider: Was the induction procedure well executed? Were appropriate control conditions utilized? Did the induction adequately operationalize the researcher's question of interest?"
Rigor of Physio Recording	1 = Low 2 = Moderate 3 = High	Categorical variable representing the rigor of the physiological recording and reporting. Coder instructions were as follows: "*Coder, this is an "inferential" variable. Many of the components that objectively make up "rigor" are recorded elsewhere. Please rate your "sense" of the methodological rigor of the physiology recording AND reporting. Some possible attributes to consider: Does physiology procedure appear well executed? Do measures reported seem appropriate given the research question? Was there an appropriate baseline procedure?*

Appendix C: Description of emotion induction procedures

Category	Description
Imagery/Autobiographical Recall	Participants engage in mental imagery to induce emotions. Most often participants are instructed to recall (and usually write down) a past event that elicited intense feelings. Autobiographical recall tasks have been used to induce happiness, sadness, anger, fear, disgust, or a neutral state. In some versions of this paradigm either a researcher or the participant reads the recall script back to the participant, either as a re-induction or as part of a different task. Physiological measures are recorded either during the writing process, during the reading of the script, or both.
Sounds and/or Music	Participants listen to a set of acoustic stimuli designed to induce emotion. Many of these paradigms use evocative music to induce emotion, but some (e.g., IADS; (Bradley, 1999) use evocative sounds like children laughing or a blaring siren. Because of the relatively small sample of physiological studies that use auditory stimuli to induce discrete emotions, we have averaged across sound and music inductions in our analysis.
Images and/or Faces	Participants view a set of static, visual stimuli designed to induce emotional responses. In these tasks, participants may view evocative images, faces, or both. The most common (non-face) stimuli are the International Affective Picture System (IAPS; (Lang, 2008), a set of normed evocative images like striking snakes or roller coasters. Evocative face stimuli are usually a set of posed, static faces displaying an expression that researchers label a "discrete" emotion (e.g., fear or disgust). In some experiments, participants are asked to mimic or imagine the mental state of the person making the face, in other experiments participants passively view the faces.
Evocative Films	Participants view video clips designed to induce emotional states. The most common of these stimulus sets induce emotional states by showing brief, highly evocative clips from popular films like *Sophie's Choice* or *The Shining* (Gross & Levenson, 1995; Philippot, 1993).
Social Evaluation	Participants perform a task (or tasks) during which they are evaluated by another person (real or implicit). In some social evaluation tasks, participants are asked to compose a speech and deliver it to a panel of experts or compute mental arithmetic while they are evaluated by a researcher (e.g., TRIER social stress task; (Kirschbaum, Pirke, & Hellhammer, 1993; Kudielka, Hellhammer, Kirschbaum, Harmon-Jones, & Winkielman, 2007).

Threat	Participants perform a task in which they are exposed to stimuli that poses a physical threat to them (real or implicit). In threat tasks, participants may be exposed to a dangerous animal or insect (e.g. participants sit in close proximity to a snake or tarantula) or they may be

Appendix D: Additional Model Parameters for Regression Models with Sample, Study, and Quality Moderators for each Emotion Category

Statistically Significant Regression Models with Regression Coefficients

	β	*F*	*df*	*p*
Anger				
Sample Moderators				
Sample Type		5.16	1, 52.4	0.03*
Student	-0.11			
Community	0.20			
ANS Measure Type x Gender		1.98	16, 156	0.02*
SBP: All Women	1.14			
SBP: Men and Women	0.84			
SBP: All Men	0.67			
DBP: All Women	0.99			
DBP: Men and Women	0.72			
DBP: All Men	0.76			
Quality Moderators				
Rigor of Physio Recording		2.10	2, 30.6	0.04*
Low Rigor	1.20			
Moderate Rigor	0.84			

High Rigor	0.97			
Fear				
Sample Moderators				
Sample Type		11.57	1, 21.3	0.003**
Student	-0.27			
Community	0.11			
ANS Measure Type x Sample Type		3.93	6, 59	0.003**
RR: Student	-0.002			
RR: Community	1.53			
HR: Student	-0.06			
HR: Community	0.83			
Happy				
Quality Moderators				
Induction Duration		6.57	3, 11	0.008**
Less than one minute	1.62			
1 – 5 minutes	-0.28			
5 – 10 minutes	-0.34			
More than 10 minutes	-0.38			
Sad				
Sample Moderators				

ANS Measure Type x Sample Type		2.60	5, 64	0.03*
RR: Student	0.22			
RR: Community	0.99			
HR: Student	0.21			
HR: Community	0.51			
Neutral				
Sample Moderators				
ANS Measure Type x Gender		2.85	10, 75	0.004**
SBP: All Women	0.47			
SBP: Men and Women	0.44			
SBP: All Men	0.13			
DBP: All Women	0.99			
DBP: Men and Women	0.11			
DBP: All Men	0.27			
SCR: All Women	0.98			
SCR: Men and Women	-0.002			
SCR: All Men	0.02			
Quality Moderators				
Manipulation Check		7.37	1, 8.45	0.03*
No manipulation check	0.22			
Manipulation check	0.04			

Note. Significant interactions between autonomic measures and moderator variables reflected the effect of the moderator on different autonomic measures. β = Regression coefficients. Regression coefficients for the categorical variables can be interpreted as the mean effect sizes for each category.
$^{\dagger}p < 0.15$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$